VERIFICATION OF TRANSLATION

I, Mariko Ogura

of 6th Fl., Kantetsu Tsukuba-Science-City Bldg. 1-1-1, Oroshi-machi, Tsuchiura, Ibaraki, JAPAN

declare as follows:

1. That I am well acquainted with both the English and Japanese languages, and

2. That the attached document is a true and correct translation made by me to the best of my knowledge and belief of:-

Japanese Patent Application No. Hei 10-214720
Entitled: "Novel Hemopoietin Receptor Proteins"
Filed on June 24, 1998

March 11, 2005
Date

Mariko Ogura
Mariko Ogura

(21000yen)

【Name of Document】 APPLICATION FOR PATENT

【Identification Number】 P10-J241

【Filing Date】 June 24, 1998

【Addressee】 Esq. Commissioner of the Patent Office

【International Patent Classification】 C12N 15/12

【Title of the Invention】 Novel Hemopoietin Receptor Proteins

【Number of Claims】 17

【Inventor】

【Address】 c/o CHUGAI RESEARCH INSTITUTE FOR MOLECULAR MEDICINE, INC.
153-2, Nagai, Niihari-mura, Niihari-gun, IBARAKI

【Name】 Hitoshi Nomura

【Inventor】

【Address】 c/o CHUGAI RESEARCH INSTITUTE FOR MOLECULAR MEDICINE, INC.
153-2, Nagai, Niihari-mura, Niihari-gun, IBARAKI

【Name】 Masatsugu Maeda

【Applicant】

【Identification Number】 596102791

【Postal Code】 300-4101

【Address】 153-2, Nagai, Niihari-mura, Niihari-gun, IBARAKI

【Name or Appellation】 CHUGAI RESEARCH INSTITUTE FOR MOLECULAR MEDICINE, INC.

【Representative】 Yoshiyuki Ohsugi

【Telephone Number】 0298-30-6211

【List of Attached Documents】

【Name of Document】 Specification 1

【Name of Document】 Drawings 1

【Name of Document】 Abstract 1

[Document Name] Specification

[Title of the Invention] NOVEL HEMOPOIETIN RECEPTOR PROTEINS

[Claims]

[Claim 1] A protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 361st amino acid Ser of SEQ ID NO: 1.

[Claim2] A protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 144th amino acid Leu of SEQ ID NO: 3.

[Claim 3] A protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 237th amino acid Ser of SEQ ID NO: 5.

[Claim 4] A protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added, and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 538th amino acid Ser of SEQ ID NO: 7.

[Claim 5] A protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 2.

[Claim 6] A protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 4.

[Claim 7] A protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 6.

[Claim 8] A protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 8.

[Claim 9] A fusion protein comprising the protein of any one of

claims 1 to 8 and another peptide or polypeptide.

[Claim 10] A DNA encoding the protein of any one of claims 1 to 9.

[Claim 11] A vector comprising the DNA of claim 10.

[Claim 12] A transformant harboring the DNA of claim 10 in an expressible manner.

[Claim 13] A method of producing the protein of any one of claims 1 to 9, comprising the step of culturing the transformant of claim 12.

[Claim 14] A method of screening a substance that binds to the protein of any one of claims 1 to 8 comprising the steps of:

(a) contacting a test sample with the protein of any one of claims 1 to 9; and

(b) selecting a substance that comprises an activity to bind to the protein of any one of claims 1 to 9.

[Claim 15] An antibody that specifically binds to the protein of any one of claims 1 to 8.

[Claim 16] A method of detecting or measuring the protein of any one of claims 1 to 9 comprising the steps of contacting a test sample presumed to contain said protein with the antibody of claim 15, and detecting or measuring the formation of the immune complex between the antibody and the protein.

[Claim 17] A DNA specifically hybridizing to a DNA comprising a nucleotide sequence of any one of SEQ ID NOs: 2, 4, 6, and 8 comprising at least 15 nucleotides, and comprising at least 15 nucleotides.

[Detailed Description of the Invention]

[0001]

[Technical Field of Industrial Application]

The present invention relates to novel hemopoietin receptor proteins, the encoding genes, and methods of production and uses thereof.

[0002]

[Prior Art]

A large number of cytokines are known as humoral factors that are involved in the proliferation/differentiation of various cells, or activation of differentiated mature cells, and also cell death. These

cytokines have their own specific receptors, which are categorized into several families based on their structural similarities (1).

Compared to similarities between receptors, primary-structure homology is quite low between cytokines, and a significant amino acid homology cannot be seen even among cytokine members that belong to the same receptor family. This explains the functional specificity of each cytokine, as well as similarities of cellular reactions induced by each cytokine.

[0003]

Representative examples of the above-mentioned receptor families are the tyrosine kinase receptor family, hemopoietin receptor family, tumor necrosis factor (TNF) receptor family, and transforming growth factor β (TGF β) receptor family. Different signal transduction pathways have been reported to be involved in each of these families. Among these receptor families, many receptors of especially the hemopoietin receptor family are expressed in blood cells and immunocytes, and their ligands, cytokines, are often termed as hemopoietic factors or interleukins. Some of these hemopoietic factors or interleukins exist within blood and are thought to be involved in a systemic humoral regulation of hemopoietic or immune functions.

[0004]

This contrasts with the belief that cytokines belonging to other families are often involved in only topical regulations. Some of these hemopoietins can be taken as hormone-like factors, and conversely, representative peptide hormones such as the growth hormone, prolactin, or leptin receptors also belong to the hemopoietin receptor family. Because of these hormone-like systemic regulatory features, it is anticipated that hemopoietin administration would be applied in the treatment of various diseases.

Among the large number of cytokines, those that are actually being clinically applied are, erythropoietin, G-CSF, GM-CSF, and IL-2. Combined with IL-11, LIF, and IL-12 that are being considered for clinical trials, and the above-mentioned peptide hormones such as growth hormone and prolactin, it can be envisaged that by searching among the above-mentioned various receptor families for a novel cytokine that binds to hemopoietin receptors, it is possible to find

a cytokine that can be clinically applied with a higher efficiency.

[0005]

As mentioned above, cytokine receptors have structural similarities between the family members. Using these similarities, many investigations are being carried out aiming at finding novel receptors. Regarding the tyrosine kinase receptor especially, many receptors have already been cloned using its highly conserved sequence at the catalytic site (2). Compared to this, hemopoietin receptors do not have a tyrosine kinase-like enzyme activity domain in their cytoplasmic regions, and their signal transductions are known to be mediated through associations with other tyrosine kinase proteins existing freely in the cytoplasm.

Though the binding site on receptors associating with these cytoplasmic tyrosine kinases (JAK kinases) is conserved between family members, the homology is not very high (3). On one hand, the sequence that characterizes these hemopoietin receptors most well exists in the extracellular region, and especially the five amino acid Trp-Ser-Xaa-Trp-Ser (where Xaa is an arbitrary amino acid) motif is conserved in almost all of the hemopoietin receptors. Therefore, novel receptors are expected to be obtained by searching novel family members using this sequence. In fact, this approach has already identified the IL-11 receptor (4), leptin receptor (5) and the IL-13 receptor (6).

[0006]

[Problems to Be Solved by the Invention]

Until now, the inventors have been trying to search for a novel receptor using an oligonucleotide encoding the Trp-Ser-Xaa-Trp-Ser motif as a probe by plaque hybridization, RT-PCR method, and so on. However, because of reasons such as the oligonucleotide tggag (t/c) nnntggag (t/c) (where n is an arbitrary nucleotide) that encodes the motif being short having just 15 nucleotides, and the g/c being high, it was extremely difficult to strictly select only those in which the 15 nucleotides have completely hybridized under the usual hybridization conditions.

[0007]

Also, a similar sequence is contained within cDNA encoding proteins

other than hemopoietin receptors, starting with various collagens that are thought to be widely distributed and also have high expression amounts, which makes the screening by the above-mentioned plaque hybridization and RT-PCR highly inefficient.

Therefore, the present invention provides a novel hemopoietin receptor protein, and the encoding DNA. The present invention also provides, a vector into which the DNA has been inserted, a transformant harboring the DNA, and a method of producing a recombinant protein using the transformant. It also provides a method of screening a substance that binds to the protein.

[0008]

[Means to Solve the Problems]

To solve these problems, and to estimate how many different hemopoietic receptor genes actually exist on the human genome, the inventors computer-searched sequences that completely coincided with each probe using all capable oligonucleotide sequences encoding the above-mentioned Trp-Ser-Xaa-Trp-Ser motif as probes.

Next, among the clones identified by the above search, the nucleotide sequence around the probe sequence of human genome-derived clones (cosmid, BAC, PAC) was converted to the amino acid sequence and compared with the amino acid sequence of known hemopoietin receptors to select genes thought to encode hemopoietin receptor family members.

[0009]

From the above search, two clones thought to be hemopoietin receptor genes were identified. One of these was the known GM-CSFβ receptor gene (derived from the 22q12.3-13.2 region of chromosome no. 22), and the other (BAC clone AC002303 derived from the 16p12 region of chromosome no. 16) was presumed to encode a novel hemopoietin receptor protein, and this gene was named "NR8."

Next, the cDNA thought to encode NR8 was found within the human fetal liver cell cDNA library by RT-PCR using a specific primer designed based on the obtained nucleotide sequence. Furthermore, using this cDNA library as the template, the full-length cDNA NR8-α encoding a transmembrane receptor comprising 361 amino acids was ultimately obtained by 5'- and 3'-RACE methods.

[0010]

In the primary structure of NR8-α, a cysteine residue and a proline rich motif conserved between other family members, were well conserved in the extracellular region, and in the intracellular region, the Box 1 motif thought to be involved in signal transduction was well conserved, and therefore, NR8α was thought to be a typical hemopoietin receptor.

[0011]

Therefore, the present invention provides:

[1] a protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 361st amino acid Ser of SEQ ID NO: 1;

[2] a protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 144th amino acid Leu of SEQ ID NO: 3;

[3] a protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 237th amino acid Ser of SEQ ID NO: 5;

[0012]

[4] a protein comprising a modified amino acid sequence of said amino acid sequence in which one or more amino acids have been deleted, added, and/or substituted with another amino acid and being functionally equivalent to the protein comprising the amino acid sequence from the 1st amino acid Met to the 538th amino acid Ser of SEQ ID NO: 7;

[5] a protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 2;

[6] a protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 4;

[7] a protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 6;

[8] a protein encoded by a DNA hybridizing to a DNA comprising the nucleotide sequence of SEQ ID NO: 8;

[9] a fusion protein comprising the protein of any one of claims 1 to 8 and another peptide or polypeptide;

[0013]

[10] a DNA encoding the protein of any one of claims 1 to 9;

[11] a vector comprising the DNA of claim 10;

[12] a transformant harboring the DNA of claim 10 in an expressible manner;

[13] a method of producing the protein of any one of claims 1 to 9, comprising the step of culturing the transformant of claim 12;

[14] a method of screening a substance that binds to the protein of any one of claims 1 to 8 comprising the steps of:

(a) contacting a test sample with the protein of any one of claims 1 to 9; and

(b) selecting a substance that comprises an activity to bind to the protein of any one of claims 1 to 9;

[0014]

[15] an antibody that specifically binds to the protein of any one of claims 1 to 8;

[16] a method of detecting or measuring the protein of any one of claims 1 to 9 comprising the steps of contacting a test sample presumed to contain said protein with the antibody of claim 15, and detecting or measuring the formation of the immune complex between the antibody and the protein; and

[17] a DNA specifically hybridizing to a DNA comprising a nucleotide sequence of any one of SEQ ID NOs: 2, 4, 6, and 8 comprising at least 15 nucleotides, and comprising at least 15 nucleotides.

[0015]

[Mode for Carrying Out the invention]

The present invention relates to the novel hemopoietin receptor proteins.

The amino acid sequences of the "NR8" proteins included in the proteins of the present invention are shown in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, and SEQ ID NO: 7, and the nucleotide sequences of cDNA encoding these proteins are shown in SEQ ID NO: 2, SEQ ID

NO: 4, SEQ ID NO: 6, and SEQ ID NO: 8, respectively. Biological activities of hemopoietin receptor proteins of the present invention are hemopoietic factor receptor protein activities.

[0016]

A cDNA encoding the protein of the invention may be obtained by, for example, screening a human cDNA library using the probe described herein.

Using the obtained cDNA or cDNA fragment as a probe, cDNA can also be obtained from other cells, tissues, organs, or species by further screening cDNA libraries. cDNA libraries may be prepared by, for example, the method of Sambrook, J. et al., Molecular Cloning, Cold Spring Harbor Laboratory Press (1989), or commercially available cDNA libraries may be used.

[0017]

By determining the nucleotide sequence of the obtained cDNA, the translation region encoded by it can be determined, and the amino acid sequence of the protein of the present invention can be obtained. Furthermore, genomic DNA can be isolated by screening the genomic DNA library using the obtained cDNA as a probe.

Specifically, this can be done as follows. First, mRNA is isolated from cells, tissues, and organs expressing the protein of the invention. For this mRNA isolation, whole RNA is prepared using well-known methods, for example, guanidine ultracentrifugation method (Chirgwin, J.M. et al., Biochemistry, 1979, 18, 5294-5299), the AGPC method (Chomczynski, P. and Sacchi, N., Anal. Biochem., 1987, 162, 156-159), and such, and purified using the mRNA Purification Kit (Pharmacia), etc. mRNA may be directly prepared using the QuickPrep mRNA Purification Kit (Pharmacia).

[0018]

cDNA is synthesized using reverse transcriptase from the obtained mRNA. cDNA can be synthesized using the AMV Reverse Transcriptase First-strand cDNA Synthesis Kit (SEIKAGAKU CORPORATION), etc. Also, cDNA synthesis and amplification may also be done using the probe described herein by following the 5'-RACE method (Frohman, M.A. et al., Proc. Natl. Acad. Sci. U.S.A., 1988, 85, 8998-9002; Belyavsky, A. et al., Nucleic Acids Res., 1989, 17, 2919-2932) using the polymerase

chain reaction (PCR) and the 5'-Ampli FINDER RACE KIT (Clontech).

[0019]

The objective DNA fragment is prepared from the obtained PCR product and ligated with vector DNA. Thus, a recombination vector is created, introduced into *E.coli*, etc. and colonies are selected to prepare the desired recombination vector. The nucleotide sequence of the objective DNA may be verified by known methods, for example, the dideoxy nucleotide chain termination method.

In the DNA of the invention, a sequence with a higher expression efficiency can be designed by considering the codon usage frequency of hosts used for the expression (Grantham, R. et al., Nucleic Acids Research, 1981, 9, r43-r74). The DNA of the invention may also be modified using commercially available kits and known methods. For example, digestion by restriction enzymes, insertion of synthetic oligonucleotides and suitable DNA fragments, addition of linkers, insertion of a start codon (ATG) and/or stop codon (ATT, TGA, or TAG), and such can be given.

The DNA of the present invention encompasses DNA comprising the nucleotide sequence from the 441st nucleotide A to the 1523rd nucleotide C in the nucleotide sequence of SEQ ID NO: 2, DNA comprising the nucleotide sequence from the 441st nucleotide A to the 872nd nucleotide A in the nucleotide sequence of SEQ ID NO: 4, DNA comprising the nucleotide sequence from the 659th nucleotide A to the 1368th nucleotide C in the nucleotide sequence of SEQ ID NO: 6, DNA comprising the nucleotide sequence from the 441st nucleotide A to the 2054th nucleotide C in the nucleotide sequence of SEQ ID NO: 8.

[0020]

The DNA of the present invention encompasses DNA that hybridizes under stringent conditions to the DNA comprising any one of the nucleotide sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, and SEQ ID NO: 8, which also includes a DNA encoding a protein having the biological activity of the protein described herein.

Stringent conditions can be suitably selected by one skilled in the art, and for example, low-stringent conditions can be given. Low-stringent conditions are, for example, 42℃, 2x SSC, and 0.1% SDS, and preferably 50℃, 2x SSC, and 0.1% SDS. More preferable are highly

stringent conditions, for example, 65℃, 2x SSC, and 0.1% SDS. Under these conditions, the higher the temperature is raised, the higher the homology of the obtained DNA will be.

The above DNA is preferably natural DNA such as cDNA and chromosomal DNA.

[0021]

As shown in Examples, the mRNA of the gene hybridizing to cDNA encoding the protein of the invention was distributed in various human tissues. Therefore, the above-mentioned natural DNA may be, for example, genomic DNA and cDNA derived from tissues in which the mRNA that hybridizes to the cDNA encoding the protein of the invention is detected in Examples. The DNA encoding the protein of the invention may be cDNA, genomic DNA, or synthetic DNA.

[0022]

To produce the protein of the invention, the obtained DNA is incorporated into an expression vector in a manner that the DNA is expressible under the regulation of an expression regulatory region, for example, an enhancer or promoter. Next, host cells are transformed by this expression vector to express the protein.

Specifically, the protein can be produced as follows. When mammalian cells are used, DNA comprising a commonly used useful promoter/enhancer, DNA encoding the protein of the invention, and the poly A signal that is functionally bound to the 3′ side downstream of the protein-encoding DNA, or a vector containing it, is constructed. For example, as the promoter/enhancer, human cytomegalovirus immediate early promoter/enhancer can be given.

[0023]

Also, as other promoters/enhancers that can be used for protein expression, viral promoters/enhancers of retroviruses, polyomaviruses, adenoviruses, simian virus 40 (SV40), and such, and promoters/enhancers derived from mammalian cells, such as that of human elongation factor 1α (HEF1α) can be used.

For example, a protein can be easily expressed by following the method of Mulligan et al. (Nature, 1979, 277, 108) when using the SV40 promoter/enhancer, and the method of Mizushima et al. (Nucleic Acids Res., 1990, 18, 5322) when using the HEF1α promoter/enhancer.

[0024]

When using *E. coli*, well-used useful promotors, the signal sequence for polypeptide secretion, and genes to be expressed, may be functionally bound to express the desired gene. For example, lacZ promoter and araB promoter may be used as promotors. When using the lacZ promoter, the method of Ward et al. (Nature, 1098, 341, 544-546; FASEB J., 1992, 6, 2422-2427), and when using the araB promoter, the method of Better et al. (Science, 1988, 240, 1041-1043) may be followed.

When producing the protein into the periplasm of *E. coli*, the pelB (Lei, S. P. et al., J. Bacteriol., 1987, 169, 4379) signal sequence may be used as a protein secretion signal.

[0025]

A replication origin derived from SV40, polyomavirus, adenovirus, bovine papilomavirus (BPV), and such may be used. To amplify gene copies in host cell lines, the expression vector may include an aminoglycoside transferase (APH) gene, thymidine kinase (TK) gene, *E.coli* xanthine guanine phosphoribosyl transferase (Ecogpt) gene, dihydrofolate reductase (dhfr) gene, and such as a selective marker.

[0026]

The expression vector used to produce the protein of the invention may be any, as long as it's an expression vector that is suitably used for the present invention. Mammalian expression vectors, for example, pEF and pCDM8; insect-derived expression vectors, for example, pBacPAK8; plant-derived expression vectors, for example, pMH1 and pMH2; animal virus-derived expression vectors, for example, pHSV, pMV, and pAdexLcw; retrovirus-derived expression vectors, for example, pZIpneo; yeast-derived expression vectors, for example, pNV11 and SP-Q01; *Bacillus subtilis*-derived expression vectors, for example, pPL608 and pKTH50; *E. coli*-derived expression vectors, for example, pQE, pGEAPP, pGEMEAPP, and pMALp2 can be given as expression vectors of this invention.

Vectors of the present invention can be used for not only producing the protein of the invention *in vivo* and *in vitro*, but also gene therapy of mammals, for example humans.

When introducing the expression vector of the present invention constructed above into a host cell, well-known methods, for example

the calcium phosphate method (Virology, 1973, 52, 456-467), electroporation (EMBO J., 1982, 1, 841-845), and such may be used.

In the present invention, an arbitrary production system may be used to produce the protein. *In vitro* and *in vivo* production systems are known as production systems for producing proteins. Production systems using eukaryotic cells and prokaryotic cells may be used as *in vitro* production systems.

[0027]

When using eukaryotic cells, production systems using, animal cells, plant cells, and fungal cells are known. As animal cells used, for example, mammalian cells such as CHO (J. Exp. Med., 1995, 108, 945), COS, myeloma, baby hamster kidney (BHK), HeLa, or Vero, amphibian cells such as *Xenopus* oocytes (Valle, et al., Nature, 1981, 291, 358-340), insect cells such as sf9, sf21, or Tn5, are known. As CHO cells, especially DHFR gene-deficient CHO cell, dhfr-CHO (Proc. Natl. Acad. Sci. USA, 1980, 77, 4216-4220), and CHO K-1 (Proc. Natl. Acad. Sci. USA, 1968, 60, 1275) can be suitably used.

[0028]

Nicotiana tabacum-derived cells are well known as plant cells, and these can be callus cultured. As fungal cells, yeasts such as the *Saccharomyces* genus, for example, *Saccharomyces cerevisiae*, filamentous bacteria such as the *Aspergillus* genus, for example, *Aspergillus niger* are known.

Bacterial cells may be used as prokaryotic production systems. As bacterial cells, *E. coli* and *Bacillus subtilis* are known.

[0029]

Proteins can be obtained by transforming these cells with the objective DNA, and culturing the transformed cells *in vitro* according to well-known methods. For example, DMEM, MEM, RPMI1640, and IMDM can be used as culture media. At that instance, fetal calf serum (FCS) and such serum supplements may be added in the above media, or a serum-free culture medium may be used. The pH is preferably about 6 to 8. Culture is usually done at about 30℃ to 40℃, for about 15 to 200 hr, and medium changes, aeration, and stirring are done as necessary.

[0030]

On the other hand, production systems using animals and plants may be given as *in vivo* production systems. The objective gene is introduced into the plant or animal, and the protein is produced within the plant or animal, and recovered. "Host" as used in the present invention encompasses such animals and plants as well. When using animals, mammalian and insect production systems can be used. As mammals, goats, pigs, sheep, mice, and cattle may be used (Vicki Glaser, SPECTRUM Biotechnology Applications, 1993). Transgenic animals may also be used when using mammals.

[0031]

For example, the objective DNA is inserted within a gene encoding a protein produced intrinsically into milk, such as goat β casein, to prepare a fusion gene. The DNA fragment containing the fusion gene is injected into a goat's embryo, and this embryo is implanted in a female goat. The protein is collected from the milk of the transgenic goats produced from the goat that received the embryo, and descendents thereof. To increase the amount of protein-containing milk produced from the transgenic goat, a suitable hormone/hormones may be given to the transgenic goats (Ebert, K.M. et al., Bio/Technology, 1994, 12, 699-702).

[0032]

Silk worms may be used as insects. When using the silk worm, it is infected with a baculovirus to which the objective DNA has been inserted, and the desired protein is obtained from the body fluids of the silk worm (Susumu, M. et al., Nature, 1985, 315, 592-594).

When using plants, for example, tobacco can be used. In the case of tobacco, the objective DNA is inserted into a plant expression vector, for example pMON 530, and this vector is introduced into a bacterium such as *Agrobacterium tumefaciens*. This bacterium is infected to tobacco, for example *Nicotiana tabacum*, to obtain the desired polypeptide from tobacco leaves (Julian, K.-C. Ma et al., Eur. J. Immunol., 1994, 24, 131-138).

[0033]

The present invention also encompasses a protein that is functionally equivalent to the protein of the present invention. Such proteins can be obtained by the method of introducing a mutation to

the amino acid sequence of a protein. For example, site-specific mutagenesis using a synthetic oligonucleotide primer, can be used to introduce a desired mutation (Kramer, W. and Fritz, H.J., Methods in Enzymol., 1987, 154, 350-367). This could also be done by a PCR-mediated site-specific mutagenesis system (GIBCO-BRL). Using these methods, the amino acid sequence of SEQ ID NO: 1, 2, 3, or 4 can be modified to obtain a protein functionally equivalent to the protein of the present invention, in which one or more amino acids in the amino acid sequence of the protein have been deleted, added, and/or substituted by another amino acid without affecting the biological activity of the protein.

[0034]

As a protein functionally equivalent to the NR8 protein of the invention, the following are given: one in which one or two or more, preferably, two to 30, more preferably, two to ten amino acids are deleted in any one of the amino acid sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, and SEQ ID NO: 7; one in which one or two or more, preferably, two to 30, more preferably, two to ten amino acids have been added into any one of the amino acid sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; or one in which one or two or more, preferably, two to 30, more preferably, two to ten amino acids have been substituted with other amino acids in any one of the amino acid sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7 as well as one which comprise any one of the amino acid sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, and SEQ ID NO: 7.

[0035]

It is already known that a protein comprising a modified amino acid sequence of a certain amino acid sequence in which one or more amino acid residues have been deleted, added, and/or substituted with another amino acid, still maintains its biological activity (Mark, D. F. et al., Proc. Natl. Acad. Sci. USA, 1984, 81, 5662-5666; Zoller, M. J. & Smith, M., Nucleic Acids Research, 1982, 10, 6487-6500; Wang, A. et al., Science, 224, 1431-1433; Dalbadie-McFarland, G. et al., Proc. Natl. Acad. Sci. USA, 1982, 79, 6409-6413).

For example, a fusion protein can be given as a protein in which

one or more amino acid residues have been added to the protein of the present invention. A fusion protein is made by fusing the NR8 protein of the present invention with another peptide or protein and is encompassed in the present invention. A fusion protein can be prepared by ligating DNA encoding the NR8 protein of the present invention with DNA encoding another peptide or protein so as the frames match, introducing this into an expression vector, and expressing the fusion gene in a host. Methods commonly known can be used for preparing such a fusion gene. There is no restriction as to the other peptide or protein that is fused to the protein of this invention.

For example, FLAG (Hopp, T.P. et al., Biotechnology, 1988, 6, 1204-1210), 6x His constituting six histidine (His) residues, 10x His, Influenza agglutinin (HA), human c-myc fragment, VSV-GP fragment, p18HIV fragment, T7-tag, HSV-tag, E-tag, SV40T antigen fragment, lck tag, α-tubulin fragment, B-tag, Protein C fragment, and such well-known peptides can be used. Examples of proteins are, glutathione-S-transferase (GST), Influenza agglutinin (HA), immunoglobulin constant region, β-galactosidase, maltose-binding protein (MBP), etc. Commercially available DNAs encoding these may also be used to prepare fusion proteins.

The protein of the invention can also be encoded by a DNA that hybridizes under stringent conditions to a DNA comprising any one of the nucleotide sequences of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, and SEQ ID NO: 8. Such a protein also includes a protein having the biological activity of the protein described herein.

[0036]

The present invention also includes a protein having the biological activity of the protein, which has also a homology with a protein comprising any one of the amino acid sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, and SEQ ID NO: 7. A protein having a homology means, a protein having at least 70%, preferably at least 80%, more preferably at least 90%, even more preferably, at least 95% homology to any one of the amino acid sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, and SEQ ID NO: 7. The homology of a protein can be determined by the algorithm in "Wilbur, W.J. and Lipman, D.J. Proc. Natl. Acad. Sci. USA, 1983, 80, 726-730."

[0037]

In the protein of the invention, the amino acid sequence, molecular weight, isoelectric point, the presence or absence of the sugar chain, and its form differ according to the producing cells, host, or purification method described below. However, as long as the obtained protein comprises an activity that is functionally equivalent to the protein of the present invention, it is included in the present invention. An activity that is functionally equivalent to a protein refers to a hemopoietic receptor protein activity that is functionally equivalent to a hemopoietic receptor protein of the present invention.

For example, if the protein of the present invention is expressed in prokaryotic cells such as *E. coli*, a methionine residue is added at the N-terminus of the amino acid sequence of the expressed protein. If the protein of the present invention is expressed in eukaryotic cells such as mammalian cells, the N-terminal signal sequence is removed. The protein of the present invention includes these proteins.

For example, as a result of analyzing the protein of the invention based on the method in "Von Heijne, G., Nucleic Acids Research, 1986, 14, 4683-4690," it was presumed that the signal sequence is from the 1st Met to the 19th Gly in the amino acid sequence of SEQ ID NO: 1. Therefore, the present invention encompasses a protein comprising the sequence from the 20th Cys to 361st Ser in the amino acid sequence of SEQ ID NO: 1.

[0038]

The present invention includes a partial peptide comprising the active center of a protein comprising any one of the amino acid sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, and SEQ ID NO: 7. A partial peptide of the protein of the present invention is, for example, a partial peptide of the molecules of the protein, which contains one or more regions of the hydrophilic region and hydrophobic region presumed by hydrophobicity plot analysis. These partial peptides may contain the whole hydrophilic region or a part of it, and may contain the whole hydrophobic region or a part of it. For example, soluble proteins and proteins comprising extracellular regions of the protein of the invention, are also encompassed in the invention.

[0039]

The partial peptides of the protein of the invention may be produced by genetic engineering techniques, well-known peptide synthesizing methods, or by excising the protein of the invention by a suitable peptidase. As peptide synthesizing methods, the solid-phase synthesizing method, and the liquid-phase synthesizing method may be used.

[0040]

The thus-obtained protein of the invention is isolated from within and without cells, or from hosts, and can be purified as a substantially pure homogenous protein. The separation and purification of the protein is not limited to any specific method and can be done using ordinary separation and purification methods used to purify proteins. For example, chromatography, filtration, ultrafiltration, salting out, solvent precipitation, solvent extraction, distillation, immunoprecipitation, SDS-polyacrylamide gel electrophoresis, isoelectric focusing, dialysis, recrystalization, and such may be suitably selected, or combined to separate/purify the protein.

[0041]

As chromatographies, for example, affinity chromatography, ion exchange chromatography, hydrophobic chromatography, gel filtration, reversed-phase chromatography, adsorption chromatography, and such can be exemplified (Strategies for Protein Purification and Characterization: A Laboratory Course Manual. Ed Daniel R. Marshak et al., Cold Spring Harbor Laboratory Press, 1996). These chromatographies can be done by liquid chromatography such as HPLC, FPLC, and the like. The present invention encompasses proteins highly purified by using such purification methods.

Proteins can be arbitrarily modified, or peptides may be partially excised by treating the proteins with appropriate modification enzymes prior to or after the purification. Trypsin, chymotrypsin, lysyl endopeptidase, protein kinase, glucosidase, and such are used as protein modification enzymes.

[0042]

The protein of the invention is useful for use in screening methods. Namely, the protein of the invention is used in the screening method that comprises the steps of contacting a test sample expected to contain

a substance that binds to the protein of the invention with the protein of the invention, and selecting the substance that comprises an activity to bind to the protein of the invention. As methods for screening a substance that comprises an activity to bind to the protein of the invention, numerous methods usually used by those skilled in the art can be employed.

[0043]

The protein of the invention that is used for these screening methods may be a recombinant, natural, or partial peptide. A substance comprising an activity to bind to the protein of the invention may be a protein comprising a binding activity, or it may be a chemically synthesized compound having a binding activity.

A protein that binds to the protein of the invention can be screened by, for example, using the West-western blotting method (Skolnik, E.Y. et al., Cell, 1991, 65, 83-90). cDNA is isolated from cells, tissues, and organs presumed to express the protein binding to the protein of the invention, this is inserted into phage vectors, for example, λgt11, ZAPII, and such, to make a cDNA library, expressed on a plate containing a culture medium, the proteins expressed are fixed on a filter, this filter is reacted with the labeled, purified protein of the invention, and plaques expressing the protein bound to the protein of the invention are detected by the labels. As methods to label the protein of the invention, the method that uses the binding ability of avidin and biotin, the method of using an antibody that specifically binds to the protein of the invention or the peptide or polypeptide fused to the protein of the invention, the method of using radioisotopes, or fluorescence, and such can be given.

[0044]

An example of screening system that provided in the present invention can screen using the two-hybrid system (Fields, S. and Sternglanz, R., Trends. Genet., 1994, 10, 286-292).

In the two-hybrid system, an expression vector containing DNA encoding the fusion protein between the protein of the invention and one subunit of a heterodimeric transcriptional regulatory factor, and an expression vector containing DNA made by ligating DNA encoding the other subunit of the heterodimeric transcriptional regulatory

factor and a desired cDNA used as a test sample are introduced into cells and expressed. If the protein encoded by the cDNA binds with the protein of the invention and the transcriptional regulatory factor forms a heterodimer, a reporter gene constructed in the cell beforehand will be expressed. Therefore, a protein binding to the protein of the invention can be selected by detecting or measuring the expression level of the reporter gene.

[0045]

Specifically, the DNA encoding the protein of the invention and the gene encoding the DNA binding domain of LexA are ligated so as the frames match to prepare an expression vector. Next, the desired cDNA and the gene encoding GAL4 transcription activation domain are ligated to prepare an expression vector.

Cells into which the HIS3 gene has been incorporated (the transcription of HIS3 gene is regulated by the promoter having a LexA binding motif) are transformed by the above two-hybrid system expression plasmids, and then incubated on a histidine-free synthetic culture medium. Herein, cells only grow when a protein interaction is present. Thus, the increase in reporter gene expression can be examined by the growth rate of the transformant.

Other than the HIS3 gene, for example, the luciferase gene, plasminogen activator inhibitor type1 (PAI-1) gene and such can be used as reporter genes.

The two-hybrid system may be constructed according to the usual methods, or a commercially available kit may be used. As commercially available two-hybrid system kits, the MATCHMARKER Two-Hybrid System, Mammalian MATCHMARKER Two-Hybrid Assay Kit (both by CLONTEC), HybriZAP Two-Hybrid Vector System (Stratagene), can be given.

[0046]

An example of screening system that provided in the present invention can screen using affinity chromatography. Namely, the protein of the invention is immobilized onto a carrier of an affinity column, and a test sample presumed to express a protein binding to the protein of the invention is applied to the column. As this test sample, a cell culture supernatant, cell extract, cell lysate, and such may be used. After applying the test sample, the column is washed to obtain

the protein binding to the protein of the invention.

[0047]

As a test sample that is used in the screening method of the present invention, for example, peptides, purified or crudely purified proteins, non-peptide compounds, synthetic compounds, microbial fermentation products, extracts of marine organisms, plant extracts, cell extracts, animal tissue extracts, and such can be given. These test samples may be novel compounds, or well-known compounds.

[0048]

The substance isolated by the screening method of the invention is a candidate drug for promoting or inhibiting the activity of the protein of the invention. The substance obtained by using the screening method of the invention encompasses a substance resulting from modifying the substance having an activity to bind to the protein of the invention by adding, deleting, and/or replacing a part of the structure.

When using the substance obtained by the screening method of the invention as drugs for humans and mammals such as, mice, rats, guinea pigs, rabbits, chicken, cats, dogs, sheep, pigs, cattle, monkeys, sacred baboons, and chimpanzees, the drug may be administered using ordinary means.

[0049]

For example, according to the need, the drugs can be taken orally as sugar-coated tablets, capsules, elixirs, and microcapsules, or parenterally in the form of injections of sterile solutions or suspensions with water or any other pharmaceutically acceptable liquid. For example, the compounds comprising the activity to bind to the protein of the invention can be mixed with physiologically acceptable carriers, flavoring agents, excipients, vehicles, preservatives, stabilizers, and binders, in a unit dose form required for generally accepted drug implementation. The amount of active ingredients in these preparations makes a suitable dosage within the indicated range acquirable.

[0050]

Examples of additives that can be mixed to tablets and capsules are, binders such as gelatin, corn starch, tragacanth gum, and arabic

gum; excipients such as crystalline cellulose; swelling agents such as cornstarch, gelatin, and alginic acid; lubricants such as magnesium stearate; sweeteners such as sucrose, lactose, or saccharin; and flavoring agents such as peppermint, Gaultheria adenothrix oil, and cherry. When the unit dosage form is a capsule, a liquid carrier, such as oil, can also be included in the above additives. Sterile compositions for injections can be formulated following usual drug implementations using vehicles such as distilled water used for injections. Active agents, naturally occurring vegetable oils such as sesame oil, palm oil can be dissolved or suspended in the vehicles.

[0051]

For example, physiological saline and isotonic liquids including glucose or other adjuvants, such as D-sorbitol, D-mannose, D-mannitol, and sodium chloride, can be used as aqueous solutions for injections. These can be used in conjunction with suitable solubilizers, such as alcohol, specifically ethanol, polyalcohols such as propylene glycol and polyethylene glycol, non-ionic surfactants, such as Polysorbate 80 (TM) and HCO-50.

[0052]

Sesame oil or soy-bean oil can be used as a oleaginous liquid and may be used in conjunction with benzyl benzoate or benzyl alcohol as a solubilizer; may be formulated with a buffer such as phosphate buffer and sodium acetate buffer; a pain-killer such as benzalkonium chloride, procaine hydrochloride; a stabilizer such as benzyl alcohol and phenol; and an anti-oxidant. The prepared injection is usually filled into a suitable ampule.

Although the dosage of the substance that has the activity to bind to the protein of the invention varies according to symptoms, the daily dose is generally about 0.1 to about 100 mg, preferably about 1.0 to about 50 mg, and more preferably about 1.0 to about 20 mg, when administered orally to an adult (body weight 60 kg).

[0053]

When given parenterally, the dose differs according to the patient, target organ, symptoms, and method of administration, but the daily dose is usually about 0.01 to about 30 mg, preferably about 0.1 to about 20 mg and more preferably about 0.1 to about 10 mg for an adult

(body weight 60 kg) when given as an intravenous injection. Also, in the case of other animals too, it is possible to administer an amount converted to 60 kg of body-weight.

[0054]

The antibody of the present invention can be obtained as a monoclonal antibody or a polyclonal antibody using well-known methods.

The antibody that specifically binds to the protein of the invention can be prepared by using the protein of the invention as a sensitizing antigen for immunization according to usual immunizing methods, fusing the obtained immunized cells with known parent cells by ordinary cell fusion methods, and screening for antibody producing cells using the usual screening techniques.

Specifically, a monoclonal or polyclonal antibody that binds to the proteins of the invention may be prepared as follows.

For example, the protein of the invention that is used as a sensitizing antigen for obtaining the antibody is not restricted by the animal species from which it is derived, but is preferably a protein derived from mammals, for example, humans, mice, or rats, especially from humans. Proteins of human origin can be obtained by using the nucleotide sequence or amino acid sequence disclosed herein.

[0055]

The protein that is used as a sensitizing antigen in the present invention can be a protein that comprises the biological activity of all the proteins described herein. Partial peptides of the proteins may also be used. As partial peptides of the proteins, for example, the amino (N) terminal fragment of the protein, and the carboxy (C) terminal fragment can be given. "Antibody" as used herein means an antibody that specifically reacts with the full-length or fragment of the protein.

[0056]

A gene encoding the protein of the invention or a fragment thereof is inserted into a well-known expression vector, and after transforming the host cells described herein, the objective protein or a fragment thereof is obtained from within and without the host cell, or from the host using well-known methods, and this protein can be used as a sensitizing antigen. Also, cells expressing the protein, cell

lysates, or chemically synthesized protein of the invention may be used as a sensitizing antigen.

The mammals that are immunized by the sensitizing antigen are not restricted, but it is preferable to select the animal by considering the adaptability with the parent cells used in cell fusion. Generally, an animal belonging to Rodentia, Lagomorpha, or Primates is used.

As animals belonging to Rodentia, for example, mice, rats, hamsters, and such are used. As animals belonging to Lagomorpha, for example rabbits, as Primates, for example monkeys, are used. As monkeys, monkeys of the infraorder Catarrhini (Old World Monkeys), for example, cynomolgus monkeys, rhesus monkeys, sacred baboons, chimpanzees, etc., are used.

[0057]

To immunize animals with the sensitizing antigen, well-known methods may be used. For example, the sensitizing antigen is generally injected into mammals intraperitoneally or subcutaneously. Specifically, the sensitizing antigen is suitably diluted, suspended in physiological saline or phosphate-buffered saline (PBS), mixed with a suitable amount of a general adjuvant if desired, for example, with Freund's complete adjuvant, emulsified and injected into the mammal. Thereafter, the sensitizing antigen suitably mixed with Freund's incomplete adjuvant is preferably given several times every four to 21 days. A suitable carrier can also be used when immunizing an animal with the sensitizing antigen. After the immunization, the elevation in the serum antibody level is detected by usual methods.

[0058]

Polyclonal antibodies against the protein of the invention can be obtained as follows. After verifying that the desired serum antibody level has been reached, blood is withdrawn from the mammal sensitized with the antigen. Serum is isolated from this blood using well-known methods. The serum containing the polyclonal antibody may be used as the polyclonal antibody, or according to needs, the polyclonal antibody-containing fraction may be further isolated from the serum.

To obtain monoclonal antibodies, after verifying that the desired serum antibody level has been reached in the mammal sensitized with the above-described antigen, immunocytes are taken from the mammal

and used for cell fusion. At this instance, immunocytes that are preferably used for cell fusion are splenocytes. As parent cells fused with the above immunocytes, preferable are mammalian myeloma cells, that have attained the feature of distinguishing fusion cells by agents.

[0059]

For the cell fusion between the above immunocytes and myeloma cells, for example, the method of Milstein et al. (Galfre, G. and Milstein, C., Methods Enzymol., 1981, 73, 3-46) is basically well known.

[0060]

The hybridoma obtained from cell fusion is selected by culturing in a usual selective culture medium, for example, HAT culture medium (hypoxanthine, aminopterin, thymidine-containing culture medium). The culture in this HAT medium is continued for a period sufficient enough for cells (non-fusion cells) other than the objective hybridoma to perish, usually from a few days to a few weeks. Next, the usual limiting dilution method is carried out, and the hybridoma producing the objective antibody is screened and cloned.

[0061]

Other than the above method of obtaining a hybridoma by immunizing an animal other than humans with the antigen, a hybridoma producing the objective human antibodies comprising the activity to bind to proteins can be obtained by the method of sensitizing human lymphocytes, for example, human lymphocytes infected with the EB virus, with proteins, protein-expressing cells, or lysates thereof *in vitro*, fusing the sensitized lymphocytes with myeloma cells derived from human, for example U266, having the capacity of permanent cell division (Unexamined Published Japanese Patent Application (JP-A) No. Sho 63-17688).

[0062]

Moreover, human antibody against the protein can be obtained using a hybridoma made by fusing myeloma cells with antibody-producing cells obtained by immunizing a transgenic animal comprising a repertoire of human antibody genes with an antigen such as a protein, protein-expressing cells, or a cell lysate thereof WO92/03918, WO93/2227, WO94/02602, WO94/25585, WO96/33735, and WO96/34096).

Other than producing antibodies by using hybridoma,

antibody-producing immunocytes such as sensitized lymphocytes that are immortalized by oncogenes may also be used.

[0063]

Such monoclonal antibodies can also be obtained as recombinant antibodies produced by using the genetic engineering technique (for example, Borrebaeck, C.A.K. and Larrick, J.W., THERAPEUTIC MONOCLONAL ANTIBODIES, Published in the United Kingdom by MACMILLAN PUBLISHERS LTD, 1990). Recombinant antibodies are produced by cloning the encoding DNA from immunocytes such as hybridoma or antibody-producing sensitized lymphocytes, incorporating this into a suitable vector, and introducing this vector into a host to produce the antibody. The present invention encompasses such recombinant antibodies as well.

[0064]

The antibody of the present invention may be an antibody fragment or a modified-antibody as long as it binds to the protein of the invention. For example, Fab, $F(ab')_2$, Fv, or single chain Fv in which the H chain Fv and the L chain Fv are suitably linked by a linker (scFv, Huston, J.S. et al., Proc. Natl. Acad. Sci. U.S.A., 1988, 85, 5879-5883) can be given as antibody fragments. Specifically, antibody fragments are produced by treating an antibody with an enzyme, for example, papain, pepsin, etc. or by constructing a gene encoding an antibody fragment, introducing this into an expression vector, and expressing this vector on suitable host cells (for example, Co, M.S. et al., J. Immunol., 1994, 152, 2968-2976; Better, M. and Horwitz, A.H., Methods Enzymol., 1989, 178, 476-496; Pluckthun, A. and Skerra, A., Methods Enzymol., 1989, 178, 497-515; Lamoyi, E., Methods Enzymol., 1986, 121, 652-663; Rousseaux, J. et al., Methods Enzymol., 1986, 121, 663-669; Bird, R.E. and Walker, B.W., Trends Biotechnol., 1991, 9, 132-137).

[0065]

As a modified antibody, an antibody bound to various molecules such as polyethylene glycol (PEG) can be used. Antibodies in the claims of the present invention encompass such modified antibodies as well. To obtain such a modified antibody, chemical modifications are done to the obtained antibody. These methods are already established in the field.

[0066]

The antibody of the invention may be obtained as a chimeric antibody comprising non-human antibody-derived variable region and a human antibody-derived constant region, or as a humanized antibody comprising non-human antibody-derived complementarity determining region (CDR), and human antibody-derived framework region (FR) and a constant region.

Antibodies thus obtained can be purified till uniform. The separation and purification methods for separating and purifying the antibody used in the present invention may be any method usually used for proteins, and is not in the least limited. Antibody concentration of the above mentioned antibody can be assayed by measuring the absorbance, or by the enzyme-linked immunosorbent assay (ELISA), etc.

[0067]

Also, as methods that assay the antigen-binding activity of the antibody of the invention, ELISA, enzyme immunoassay (EIA), radio immunoassay (RIA), or fluorescent antibody method can be given. For example, when using ELISA, the protein of the invention is added to a plate coated with the antibody of the invention, and next, the objective antibody sample, for example, culture supernatants of antibody-producing cells, or purified antibodies are added. Then, secondary antibody recognizing the antibody, which is labeled by alkaline phosphatase and such enzymes, is added, the plate is incubated and washed, and absorbance is measured to evaluate the antigen-binding activity after adding an enzyme substrate such as p-nitrophenyl phosphate. As the protein, a protein fragment, for example, a fragment comprising a C terminus, or a fragment comprising an N terminus may be used. To evaluate the activity of the antibody of the invention, BIAcore (Pharmacia) may be used.

[0068]

By using these methods, the antibody of the invention and a sample presumed to contain the protein of the invention are contacted, and the protein of the invention is detected or assayed by detecting or assaying the immune complex of the above-mentioned antibody and protein.

A method of detecting or assaying the protein of the invention is useful in various experiments using proteins as it can specifically

detect or assay the proteins.

[0069]

The present invention also encompasses a DNA specifically hybridizing to a DNA comprising a nucleotide sequence of any one of SEQ ID NOs: 2, 4, 6, and 8 or its complementary DNA, and comprising at least 15 nucleotides. Namely, a probe that can selectively hybridize to the DNA encoding the protein of the invention, or a DNA complementary to the above DNA, a nucleotide or nucleotide derivative, for example, antisense oligonucleotide, ribozyme, and such are included.

[0070]

The present invention also encompasses an antisense oligonucleotide that hybridizes to any portion of any one of the nucleotide sequences shown in, for example, SEQ ID NOs: 2, 4, 6, and 8. This antisense oligonucleotide is preferably one against at least 15 continuous nucleotides in any one of the nucleotide sequences of SEQ ID NOs: 2, 4, 6, and 8. More preferable is the above-mentioned antisense oligonucleotide against the above-mentioned at least 15 continuous nucleotides containing a translation start codon.

[0071]

Derivatives or modified products of antisense oligonucleotides can be used as antisense oligonucleotides. As such modified products, for example, lower alkyl phosphonate modifications such as methyl-phosphonate-type or ethyl-phosphonate-type, phosphorothioate or phosphoroamidate-modified products, etc. may be used.

[0072]

The term "antisense oligonucleotide(s)" as used herein means, not only those in which the nucleotides corresponding to those constituting a specified region of a DNA or mRNA are entirely complementary, but also those having a mismatch of one or more nucleotides, as long as the DNA or mRNA and the oligonucleotide can selectively and stably hybridize with the nucleotide sequence of SEQ ID NO: 1.

"Selectively and stably hybridize" means that significant cross hybridization with DNA encoding other proteins does not occur under usual hybridization conditions, preferably under stringent

hybridization conditions. Such DNAs are indicated as those having, in the "at least 15 continuous nucleotide" sequence region, a homology of at least 70% or higher, preferably 80% or higher, more preferably 90% or higher, even more preferably 95% or higher nucleotide sequence homology. The algorithm stated herein can be used to determine homology. Such DNA is useful as a probe for detecting or isolating DNA encoding the protein of the invention, or as a primer for amplification as described in Examples below.

[0073]

The antisense oligonucleotide derivative of the present invention acts upon cells producing the protein of the invention by binding to the DNA or mRNA encoding the protein to inhibit its transcription or translation, and to promote the degradation of mRNA, and has an effect of suppressing the function of the protein of the invention by suppressing the expression of the protein.

[0074]

The antisense oligonucleotide derivative of the present invention can be made into an external preparation such as a liniment and a poultice by mixing with a suitable base material, which is inactive against the derivatives.

Also, as needed, the derivatives can be formulated into tablets, powders, granules, capsules, liposome capsules, injections, solutions, nose-drops, and freeze-dried agents by adding excipients, isotonic agents, solubilizers, stabilizers, preservatives, pain-killers, etc. These can be prepared using the usual methods.

The antisense oligonucleotide derivative is given to the patient by directly applying onto the ailing site, by injecting into a blood vessel, etc. so that it will reach the ailing site. An antisense-mounting material can also be used to increase durability and membrane-permeability. Examples are, liposome, poly-L lysine, lipid, cholesterol, lipofectin, or derivatives of these.

[0075]

The dosage of the antisense oligonucleotide derivative of the present invention can be adjusted suitably according to the patient's condition and used in desired amounts. For example, a dose range of 0.1 to 100 mg/kg, preferably 0.1 to 50 mg/kg can be administered.

The antisense oligonucleotide derivative of the present invention is useful in inhibiting the expression of the protein of the invention, and therefore is useful in suppressing the biological activity of the protein of the invention. Also, expression-inhibitors comprising the antisense oligonucleotide derivative of the present invention are useful because of their capability to suppress the biological activity of the protein of the invention.

[0076]

[Examples]

The present invention shall be described in detail below with reference to examples, but is not be construed as being limited thereto.

[0077]

Materials and methods

1) Two step Blast Search

Probe sequences (256 types) comprising the tggag(t/c)nnntggag(t/c) (where n is an arbitrary nucleotide) as the oligonucleotide encoding the Trp-Ser-Xaa-Trp-Ser motif were designed. These sequences enable the detection of almost all known hemopoietin receptors, except for the EPO receptor, TPO receptor, and the mouse IL6 receptor. Using each sequence as the query, the GenBank nr database was searched using the BlastN (Advanced BlastN 2.0.4) program. Default values (Descriptions=100, Alignments=100) were used as parameters for the search, except for making the expectation value 100.

[0078]

Since approximately 500 clones that completely matched the probe sequences were obtained as a result of the primary search, among these, a 180-residue nucleotide sequence of human genome-derived clones (cosmid, BAC, and PAC) containing the probe sequence in approximately the center was excised. Next, using this 180-residue nucleotide sequence as the query, the nr database was searched again using the BlastX (Advanced BlastX 2.0.4) program to search the homology of the amino acid sequence around the probe sequence with known hemopoietin receptors.

Default values were used as parameters for the search, except for making the expectation value 100. However, when extremely large number of hits were obtained (caused by the Alu sub family that is a high

repetitive sequence), it was often difficult to observe hits for known hemopoietic receptors. Therefore, to maximize the sensitivity in such cases, a value of Expect=1000, Descriptions=500, Alignments=500 was used.

[0079]

For each clone that hit one or more known hemopoietin receptors as a result of the secondary search, further investigation was done to confirm that the hit matched the reading frame for the Try-Ser-Xaa-Trp-Ser motif, and there was no inframe stop codon within the query sequence. Clones that did not match the above-described search conditions were excluded. It should be noted that the validity of the above-described search conditions has been previously verified using known hemopoietin receptors, the EPO receptor and the G-CSF receptor, as positive controls.

Furthermore, to search an exon adjacent to the exon containing the Trp-Ser-Xaa-Trp-Ser motif, a BlastX search was done under the above-described conditions, by excising a sequential 180-residue nucleotide sequence in both the 5′ and 3′ directions, centering on the query sequence used in the secondary search, and using it as a query. This search detected additional partial exon sequences in both the 5′ and 3′ sides. The sequences thus obtained were used to design primers for RT-PCR as described in the next section.

[0080]

2) Search for NR8 expressing tissues using RT-PCR

To identify NR8 expressing tissues, in the AC002303 sequence of the above-described BAC clone, several exon regions widely conserved at the amino acid translation level in known cytokine receptors were surmised, and on the sequence of the surmised exon region, the following primers were synthesized. (See Fig. 5 for the location of each primer.)

NR8-SN1; 5′- CCG GCT CCC CCT TTC AAC GTG ACT GTG ACC -3′ (SEQ ID NO: 9)

NR8-SN2; 5′- GGC AAG CTT CAG TAT GAG CTG CAG TAC AGG -3′ (SEQ ID NO: 10)

NR8-AS1; 5′- ACC CTC TGA CTG GGT CTG AAA GAT GAC CGG -3′ (SEQ ID NO: 11)

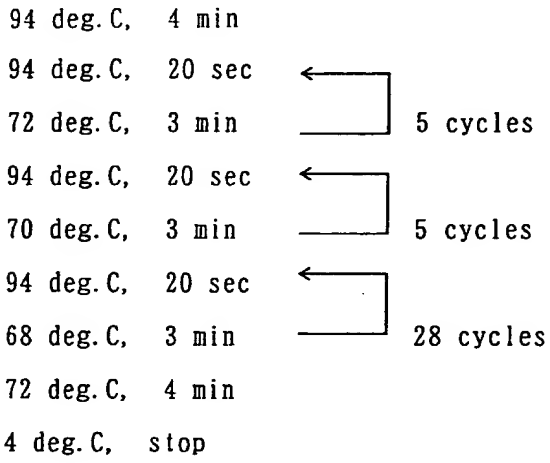
NR8-AS2; 5′- CAT GGG CCC TGC CCG CAC CTG CAG CTC ATA -3′ (SEQ ID NO:

12)

[0081]

Using the Human Fetal Multiple Tissue cDNA Panel (Clontech #K1425-1) as the template, RT-PCR was attempted using combinations of the above primers. Advantage cDNA Polymerase Mix (Clontech #8417-1) was used for the PCR, which was conducted under the conditions below using the Perkin Elmer Gene Amp PCR System 2400 Thermalcycler.

The PCR condition



[0082]

The obtained PCR product was subcloned to pGEM-T Easy vector (Promega #A1360), and the nucleotide sequence was determined. The recombination of PCR products to the pGEM-T Easy vector was done by T4 DNA Ligase (Promega #A1360) reacted at 4℃ for 12 hr. The genetic recombinant between the PCR product and pGEM-T Easy vector was obtained by transforming *E.coli* strain DH5α (Toyobo #DNA-903).

For the selection of the genetic recombinant, Insert Check Ready (Toyobo #PIK-101) was used. The dRhodamine Terminator Cycle Sequencing Kit (ABI/Perkin Elmer #4303141) was used for determining the nucleotide sequence, and analysis was done using the ABI PRISM 377 DNA Sequencer. As a result of determining the nucleotide sequences of all inserts of the 10 independent clones of genetic recombinants, all clones were found to comprise a single nucleotide sequence. These obtained sequences were verified to be partial nucleotide sequences of NR8.

[0083]

3) Full-length cDNA cloning by the 5′ and 3′-RACE methods

Using the thus-obtained fetal liver-derived cDNA, 5′ and 3′-RACE methods were conducted to obtain full-length cDNA (Fig. 4).

3-1) 5′-RACE method

To isolate full-length NR8 cDNA, 5′-RACE PCR was performed using the above-mentioned NR8-AS1 primer for primary PCR, and NR8-AS2 primer for secondary PCR. Human Fetal Liver Marathon-Ready cDNA Library (Clontech #7403-1) was used as the template and Advantage cDNA Polymerase Mix for the PCR experiment. As a result of PCR under the following conditions using the Perkin Elmer Gene Amp PCR System 2400 Thermalcycler, two types of PCR products were obtained, which have different sizes through selective splicing.

1.st PCR	94 deg.C,	4 min	
	94 deg.C,	20 sec	←
	72 deg.C,	4 min	5 cycles
	94 deg.C,	20 sec	←
	70 deg.C,	4 min	5 cycles
	94 deg.C,	20 sec	←
	68 deg.C,	4 min	28 cycles
	72 deg.C,	4 min	
	4 deg.C,	stop	

2.nd PCR	94 deg.C,	4 min	
	94 deg.C,	20 sec	←
	70 deg.C,	3 min 30 sec	5 cycles
	94 deg.C,	20 sec	←
	68 deg.C,	3 min 30 sec	28 cycles
	72 deg.C,	4 min	
	4 deg.C,	stop	

[0084]

Both types of PCR products obtained were subcloned to pGEM-T Easy vector as mentioned earlier, and the nucleotide sequences of all inserts were determined for the 16 independent clones of genetic transformants.

As before, the dRhodamine Terminator Cycle Sequencing Kit was used for determining the nucleotide sequence, and analysis was done using the ABI PRISM 377 DNA Sequencer. As a result, the clones can be divided into two groups, one having 14 clones, and the other having 2 clones, by the length of the base pairs and the differences in sequence (though described later, the differences lie in the products due to selective splicing, and the group of 14 independent clones comprises the sequence corresponding to exon 5 in the genomic sequence, and the remaining group of two independent clones does not have this sequence).

[0085]

3-2) 3'-RACE method

To isolate full-length NR8 cDNA, 3'-RACE PCR was performed using the above-mentioned NR8-SN1 primer for primary PCR, and NR8-SN2 primer for secondary PCR. Human Fetal Liver Marathon-Ready cDNA Library was used as the template similar to 5'-RACE PCR, and Advantage cDNA Polymerase Mix for the PCR experiment. As a result of conducting PCR under the conditions shown in 3-1), a single band PCR product was obtained.

The obtained PCR product was subcloned to pGEM-T Easy vector as above, and the nucleotide sequences of all inserts of the 12 independent clones of genetic recombinants were determined. As before, the dRhodamine Terminator Cycle Sequencing Kit was used for determining the nucleotide sequence, and the sequences determined were analyzed using the ABI PRISM 377 DNA Sequencer. As a result, all 12 independent clones showed a single nucleotide sequence. A nucleotide sequence determined from the result of 3'RACE-PCR and a nucleotide sequence determined from the result of 5'RACE-PCR described above were combined to determine a nucleotide sequence of full-length NR8 cDNA.

[0086]

4) Northern blotting

In order to analyze the distribution and mode of NR8 gene expression in each human organ and human cancer cell lines, Northern blot analysis was done using the cDNA clones obtained by PCR described above as a probe. The probe was prepared using Mega Prime Kit (Amersham, cat#RPN1607) and radiolabeled with [α-^{32}P] dCTP (Amersham, cat#AA0005). A probe fragment of 5'RACE-PCR product and a probe fragment of 3'RACE-PCR

product were mixed with molar ratio 1:1.

As Northern blots, Human Multiple Tissue Northern (MTN) Blot (Clontech #7760-1), Human MTN Blot IV (Clontech #7766-1), and Human Cancer Cell Line MTN Blot (Clontech #7757-1) were used. Express Hyb Hybridization Solution (Clontech #8015-2) was used for hybridization.

[0087]

Hybridization conditions were: a prehybridization at 68℃ for 30 min, followed by hybridization at 68℃ for 14 hr. After washing under the following conditions, the blots were exposed to Imaging Plate (FUJI#BAS-III), and the gene expression of NR8 mRNA was detected by the Image Analyzer (FUJIX, BAS-2000 II).

Washing condition

(1) 1x SSC/0.1% SDS, at room temperature for 5 min

(2) 1x SSC/0.1% SDS, at 50℃ 30 min

(3) 0.1x SSC/0.1% SDS, at 50℃ 30 min

Results

[0088]

About 500 hits were obtained by BlastN search using 256 probe sequence as the query (May 30, 1998). Clones derived from human accounted for about one third of the hits. Twenty-eight clones hit one or more known hemopoietin receptors (Table 1).

[Table 1]

Four clones out of these 28 clones (AC002303, AC003112, AL008637, and AC004004) hit several known hemopoietin receptors, however, AC004004 was excluded as it has a stop codon downstream three amino acids of the Trp-Ser-Xaa-Trp-Ser motif. Among the three remaining clones, AL008637 was thought to be a known receptor, GM-CSF receptor β. AC002303 is the BAC clone CIT987-SKA-670B5 derived from the 16p12 region of human chromosome no. 16 registered by TIGR group on June 19, 1997 and comprises the full-length of 131530 base pairs.

[0088]

As shown in Fig. 1, a BlastX search (query: 180 nucleotides of 40861-41040 including tggagtgaatggagt (40952-40966), the only probe sequence within the AC002303) revealed that numerous hemopoietin receptors starting with the TPO receptor and leptin receptor show an evident homology, however, there were no known, database-registered

hemopoietin receptors that completely matched the query sequence. Also, a BlastX scanning was done, by excising a sequential 180-residue nucleotide sequence in both the 5′ and 3′ directions, centering on the 180-residue nucleotide sequence mentioned above, and when this was used as a query, two sequences having a homology to known hemopoietin receptors were found in the regions 39181-39360 and 42301-42480, and were thought to be other exons of the same gene (Fig. 2).

A Pro-rich motif PAPPF was conserved in the 39181-39360 site, and a Box 1 motif in the 42301-42480 site. The 3′ side exon adjacent to the exon containing the Trp-Ser-Xaa-Trp-Ser motif has a transmembrane domain, and this domain has a low homology with other hemopoietin receptors, and was not detected by the BlastX scan. These results suggested the possibility of a novel hemopoietin receptor gene existing in the above-described BAC clone CIT987-SKA-670B5.

[0089]

Pseudogenes have been reported to exist in several hemopoietin receptors. To verify that NR8 is not a pseudogene, transcripts of the NR8 gene were searched by RT-PCR method. From the primer locations shown in Fig. 5, amplifications of bands sized 330 bp, 258 bp, 234 bp, and 162 bp can be expected from the combinations of SN1/AS1, SN1/AS2, SN2/AS1, and SN2/AS2. When evaluated using human fetal liver, brain, and skeletal muscle cDNA as the template, clear bands having the anticipated sizes were obtained in the fetal liver only with the respective primer combinations (Fig. 3).

An amplification was not seen at all for fetal brain cDNA, and a band of about 650 bp and a broad band of 400 to 500 bp were observed for fetal skeletal muscle cDNA. However, since the band sizes for skeletal muscle cDNA remained constant even when different combinations of primers were used, it is thought that these bands were non-specific amplifications due to some reason. Using the thus-obtained fetal liver-derived cDNA, 5′ and 3′-RACE methods were conducted to obtain full-length cDNA (Fig. 4).

[0090]

As a result of analyzing the nucleotide sequence of the fragments (approximately 1.1 kb and 1.2 kb) amplified by 5′-RACE and 3′-RACE, respectively, it was conceived that the approximately 260 bp of each

fragment overlap and extend to the 5′ side and 3′ side, and contain almost the full-length of NR8 mRNA. These were joined to make a full-length cDNA (NR8α) (Fig. 5).

As shown in Fig. 5, in the ORF of NR8α cDNA, the Met starting from nucleotide no. 441 is thought to be the start codon due to the presence of an inframe stop codon 39 bp upstream, and completes with two stop codons starting from nucleotide no. 1524. It has the features of, from the N terminus in order, a typical secretion signal sequence, a domain thought to be the ligand binding site containing a Cys residue conserved in other hemopoietic receptor members, a Pro-rich motif, Trp-Ser-Xaa-Trp-Ser motif, a transmembrane domain, a Box 1 motif thought to be involved in signal transduction, and such features of hemopoietin receptors. From the above results, the NR8 gene was thought to encode a novel hemopoietin receptor.

[0091]

Analysis of fragments amplified by the RACE method suggested the presence of a splice variant. As a result of nucleotide sequence analysis, this variant was revealed to be lacking approximately 150 bp including the above-described Pro-rich motif of NR8α. Moreover, as a result of comparing AC002303 sequence with NR8α, and carrying out analogy of exons/introns (Table 2), the above-described variant was thought to be deficient of the 5^{th} exon due to alternative splicing. [Table 2]

This variant (NR8β) can encode a soluble receptor in the truncated form by the joining of the 6^{th} exon directly to the 4^{th} exon and causing a frame shift. The boundary between the exons and the introns takes a consensus sequence in most cases, but the boundary between the 9^{th} exon (Exon 9a) and the 9^{th} intron is the only boundary that takes a different sequence from the consensus sequence (nag/gtgagt, etc.), being acc/acggag. Thus, it is possible theoretically to predict a potential sequence (exon 9b) based on the assumption that no splicing occurs at this site, although there is no evidence of the presence of mRNA coincident with the sequence in the present examination. Accordingly, such hypothetical sequence was named NR8γ. NR8γ encodes a protein that contains an insertion of 177 amino acids around the c-terminus of NR8α. In addition, Figs. 6 and 7 show cDNA sequences

of NR8α and NR8γ, respectively.

[0092]

Fig. 8 shows the results of Northern blot analysis of the NR8 expression in various organs. Different sized mRNAs were detected in human adult lung, spleen, thymus, skeletal muscle, pancreas, small intestine, peripheral leucocytes, and uterus. A similar examination in various cell lines including hematopoietic cell lines also showed the expression patterns in two cell lines, the promyeloid leukemic cell line HL60 and the Burkitt's lymphoma-derived Raji.

Total of three different sized bands, one 5 kb-sized and two 3 to 4 kb-sized, were observed in spleen, thymus, peripheral leucocytes, lung, and the above leukemic cell line. On the other hand, a 2 kb-sized mRNA in skeletal muscle, small intestine, and uterus, and a 1.2 kb-sized mRNA in pancreas, both of which are small, thought to be either degradation or non-specific cross reaction products.

[0093]

DISCUSSION

The two-step Blast search identified a human genomic sequence containing a novel hemopoietin receptor gene. In the present examination, the primary search was done manually using 256 types of 15-residue oligonucleotide sequences encoding all possible Trp-Ser-Xaa-Trp-Ser sequences as the query. In the preliminary examination, a tBlastN search was employed using amino acid sequence as a query, in order to save time needed for the query, but no hit was obtained even at the highest level of sensitivity, when a conservative sequence of known receptors having 5 amino acids was used as a query.

When an extended Trp-Ser-Xaa-Trp-Ser motif to both the 5' and 3' ends was used as the query, it was found that at least 8 amino acids in length was needed to obtain the hit. The inventors thought that oligonucleotide sequence might be preferable as the query, since all possible number of sequences having 8 amino acids including the Trp-Ser-Xaa-Trp-Ser motif may be 20^4, and the amino acid query may also hit a sequence containing TCN used as a codon for the serine residue. The primary search corresponds to plaque hybridization done on a computer or *in silico* using a degenerate probe, which may

necessarily hit many pseudo-positive clones.

[0094]

In the present examination, approximately 500 primary hits were obtained in despite of selecting only the hit completely matched the 15-residue sequence. Representative examples of the pseudo-positive clones include genes for thrombospondin, collagen, semaphorin, *Alu* sequence, reverse transcriptase, components of complement, Notch, and the T cell receptor. Some of these clones were also obtained frequently as the pseudo-positive clones in the actual plaque hybridization. The primary 500 hits contained almost all known hemopoietin receptor cDNAs except for the EPO receptor, TPO receptor, and the mouse IL-6 receptor.

Approximately one third of the primary 500 hits, 157 clones (including 14 overlapping clones), were derived from human genomic clones (cosmid, BAC, and PAC), which distributed in all chromosomes except for chromosomes 2, 8, and 10 (Table 1). Since these genomic sequences have not been completely analyzed yet due to their complexity at the present, they were thought to be useful as materials for screening unknown receptors.

[0095]

Also, if a total number of the genomic sequences registered to the database and a total number of nucleotides included in the registration are known at the search point, then it is possible to predict a total number of the sequences encoding the Trp-Ser-Xaa-Trp-Ser motif present on the human chromosome, as well as a total number of the hemopoietin receptors including unknown receptors. Assuming that 5% of the total genome is covered with the human genome sequences on the nr database, then 60 hemopoietin receptors per total genome are estimated to exist since three hemopoietin receptor genes were detected in the present examination.

The secondary search using the BlastX is equivalent to a homology search carried out in order to judge whether the candidate clones obtained by plaque hybridization encode the hemopoietin receptors. The reason why a 180-residue nucleotide was excised and used it as a query, is based on the findings that the general size of the exon in the case of known hemopoietin receptor genes containing the

Trp-Ser-Xaa-Trp-Ser motif is around 180 nucleotides in length, and that the GenBank report in the manual search is provided in a format of 60 nucleotides per line, which is convenient.

[0096]

The facts that the Trp-Ser-Xaa-Trp-Ser motif is located in 3' side a little apart from center of the exon and the length of the exon is a little different among the hemopoietin receptor genes, as shown in Figs. 1 and 2, indicate a portion of the query sequence includes intron. However, it is clear that the presence of the intron sequence does not interfere with the search, since, in fact, three hemopoietin receptor genes and one genomic sequence thought to be a pseudogene were detected with the sensitivity used in this examination.

A schematic representation of the NR8 gene structure, as shown in Fig. 9, indicates that the region in which the NR8 gene is located is almost filled with repetitive sequences including *Alu* subfamily and MIR, and the NR8 exons are scattered over very limited gaps within these repetitive sequences. Among the 60 repetitive sequences distributed in this region, no overlapping sequence is observed between the exons and these repetitive sequences except that only the (CA)n repeat is present in the 3'-UTR of the 10^{th} exon of the NR8 gene (Fig. 9).

Also, it is well conceivable that the presence of above highly repetitive sequences around the NR8 gene inhibited the detection of the exon of the NR8 gene using the Grail program or the detection of a homology to known hemopoietin receptor genes. It can be said that the short quick step search using the short query sequence such as 180 nucleotides lead to the detection of the exon surrounded with these repetitive sequences.

[0097]

If a similar search was done using a longer sequence as the query, many sequences might have hit the flanking repetitive sequences. As a result, it might be difficult to detect the sequence containing the Trp-Ser-Xaa-Trp-Ser motif. The method of the present examination may be useful for the detection of a gene adjacent to the highly repetitive sequences, like the NR8 gene. Although biological significances of these repetitive sequences are not known, it is

possible to suggest that the presence of the repetitive sequences reduce the gene stability.

Whereas only the sequences in which the 15 nucleotides were completely identical with the query sequence were selected in the primary search of the present examination, it should be noted that there exists known hemopoietin receptors with partially irregular forms of the motif (IL-3 receptorα, mouse IL-2 receptor p40, growth hormone receptor, and such).

[0098]

Also, as described before, EPO receptor, TPO receptor, and IL-6 receptor are excluded from the search target because the second serine of the motif is encoded by the TCN codon in the case of these receptors. In particular, when a preliminary search was done using a sequence having the TCN codon for the second serine as a query, many hits were obtained against immunoglobulin-like receptors which are homologous to IL-6 receptor (many reports appeared in 1997[10, 11]).

In addition to the Trp-Ser-Xaa-Trp-Ser motif corresponding to the query sequence, there exists another (Val/Leu)-Glu-Leu-(Val/Leu)-Val motif in the different frame, which form a large family consisting many members. It may be possible to expect that there exists a useful receptor among sequences excluded from the above consensus sequence, whose search was not done in the present examination.

[0099]

At least three different cDNAs, NR8α, NR8β, and NR8γ were expected to exist from the results of both the 5′- and 3′-RACE analyses and the genomic sequence analysis of the NR8 gene. Among them, NR8β is a alternatively spliced product lacking the 5th exon, which is possible to encode two different proteins, one is a soluble protein in which the CDS terminates at the stop codon generated on the 6th exon by a frame shift caused by direct binding of the 6th exon to the 4th exon, and the other is a membrane-binding protein lacking a signal sequence, in which the CDS starts at the ATG codon on the 4th exon.

Between them, the soluble protein has the same amino acid sequence as that of NR8α started from the first amino acid to the sequence encoded by the 4th exon, suggesting that it functions as a soluble receptor. On the other hand, NR8γ is a potential transcript in which

the CDS is read through the 9th intron and connects in frame with the 10th exon, on the assumption that the splicing does not occur between the 9th exon and the 9th intron of the NR8α gene based on the difference in the boundary sequence between the 9th exon and the 9th intron from the consensus sequence.

[0100]

As a result, the 9th exon of the NR8γ gene extends to approximately 1100 bp in length and contains a 177 amino acid insertion around the C terminus of NR8α. Both the NR8α and NR8γ genes encode transmembrane type hemopoietin receptors. Intracellular domains of both NR8α and NR8γ contain a Box1-like motif near at the cell membrane, which is one of conservative sequences among other hemopoietin receptors and thought to be involved in signal transduction. A Box2-like sequence also exists, though the conservation levels are low, suggesting that NR8 belongs to such a receptor that mediates signal transduction as a homodimer.

It may be possible to confirm that whether NR8 can actually transduce a ligand-dependent proliferation signal, and which receptor has the activity between NR8α and NR8γ in that case, by constructing a chimeric receptor generated by a fusion of each intracellular domain and a extracellular domain of a known hemopoietin receptor and examining a growth-stimulating activity of a hemopoietic factor-dependent cell line in which the chimeric receptor has been introduced, by stimulating with the known hemopoietin.

[0101]

As a result of Northern blot analysis, multiple bands were detected at the positions approximately 5 kb, 3-4 kb, 2 kb, and 1.2 kb in various tissues and cell lines. Among them, the 2 kb band was observed only in skeletal muscle, small intestine, and uterus. On the other hand, as a result of RT-PCR, an amplified band was detected in fetal skeletal muscle (Fig. 3).

However, whereas different-sized bands were detected in the same RT-PCR as expected when different-positioned primers and cDNA derived from fetal liver as a template were used, in contrast, no difference in the size of the amplified bands was observed in fetal skeletal muscle using the same primer set. Thus, the amplified fragment in

skeletal muscle may be different from that of the NR8 gene, suggesting the presence of a transcript homologous to NR8 in skeletal muscle. Probably, observation of the 2 kb band in above skeletal muscle was a result of cross hybridization of the probe to the transcript homologous to NR8, which may also explain the 2 kb band observed in both small intestine and uterus.

[0102]

On the other hand, the 1.2 kb short transcript was detected only in pancreas among the tissues examined. This band is not expected to be the transcript of the NR8 gene because the transcriptional initiation site of NR8α is predicted to exist upstream from 1th nucleotide shown in Fig. 5 and therefore the size of the NR8 mRNA is estimated to be longer than 1884 nucleotides without poly A tail. Though the possibility that the 1.2 kb band is a degradation product cannot be rule out since pancreas is an organ rich in many kind of hydrolysing enzymes, observation of the 1.2 kb band was probably resulted from the same cross hybridization as the cases in skeletal muscle, small intestine, and uterus.

Two to three bands were observed in the 5 kb and 3-4 kb regions in other tissues than those mentioned above (spleen, thymus, peripheral leucocytes, and lung). Similar-sized bands were also detected in cell lines HL60 and Raji, but no expression was observed in other cancer cell lines (HeLa, SW480, A549, and G631) and in leukemia cell lines (K562 and MOLT4).

These results suggest that NR8 is expressed specifically in hemopoietic cells, particularly in granular cells and B cells. The size of the full-length NR8 mRNA including 5'- and 3'-UTR has not been estimated yet from the results of 5'- and 3'-RACE analyses. Probably, above-described different transcripts reflect different sized NR8 transcripts including these UTRs, and the different sized transcripts correspond to the splice variants.

[0103]

As for the medical application of the NR8 protein, first of all, NR8 is suggested to be a receptor for an unknown hemopoietic factor by the fact that it is expressed in fetal liver, spleen, thymus, and a kind of leukemia cell line. Therefore, the NR8 protein may be a

useful material to obtain the unknown hemopoietic factor.

Moreover, since NR8 is expected to be expressed specifically in a limited cell population within these hemopoietic tissues, an anti-NR8 antibody may be useful for the isolation of the cell population. The cell population thus obtained may be applied to a cell transplantation therapy. Furthermore, the anti-NR8 antibody is expected to be applied to typing or therapy for diseases including leukemia. On the other hand, the soluble protein containing the extracellular domain of the NR8 protein or the NR8β protein as the splice variants of NR8 is expected to be used for an inhibitor of a NR8 ligand as a decoy type receptor, and expected to be applied to therapy for diseases including leukemia in which NR8 involves.

[0104]

References

1) Hilton D.J., in "Guidebook to Cytokines and Their Receptors" edited by Nicola N.A. (A Sambrook & Tooze Publication at Oxford University Press), 1994, p8-16

2) Matthews W. et al., Cell, 1991, 65 (7) p1143-52

3) Murakami M. et al., Proc. Natl. Acad. Sci. USA, 1991, 88, 11349-11353

4) Robb, L. et al., J. Biol. Chem., 1996, 271 (23) 13754-13761

5) Gainsford T. et al., Proc. Natl. Acad. Sci. USA, 1996, 93 (25) p14564-8

6) Hilton D.J. et al., Proc. Natl. Acad. Sci. USA, 1996, 93 (1) p497-501

7) Kermouni, A. et al., Genomics, 1995, 29 (2) 371-382

8) Fukunaga, R. and Nagata, S., Eur. J. Biochem., 1994, 220, 881-891

9) Lamerdin, J.E., et al., GenBank Report on AC003112, 1997

10) Cosman, D., et al., Immunity, 1997, In press

11) Wagtmann, N., et al., Curr. Biol. 7 (8), 1997, 615-618

[0105]

[Effects of the Invention]

The present invention provides a novel hemopoietin receptor protein, and the encoding DNA. The present invention also provides, a vector into which the DNA has been inserted, a transformant harboring the DNA, and a method of producing a recombinant protein using the transformant. It also provides a method of screening a substance that binds to the protein. The protein of the invention is thought to be

related to hemopoiesis, and therefore, is useful in experiments for analyzing hemopoietic functions.

[0106]

[Sequence Listing]

SEQ ID NO: 1

SEQUENCE LENGTH: 361

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

SEQUENCE TYPE: protein

SEQUENCE DESCRIPTION

Met Pro Arg Gly Trp Ala Ala Pro Leu Leu Leu
1 5 10
Leu Leu Leu Gln Gly Gly Trp Gly Cys Pro Asp Leu Val Cys Tyr Thr
15 20 25
Asp Tyr Leu Gln Thr Val Ile Cys Ile Leu Glu Met Trp Asn Leu His
30 35 40
Pro Ser Thr Leu Thr Leu Thr Trp Gln Asp Gln Tyr Glu Glu Leu Lys
45 50 55
Asp Glu Ala Thr Ser Cys Ser Leu His Arg Ser Ala His Asn Ala Thr
60 65 70 75
His Ala Thr Tyr Thr Cys His Met Asp Val Phe His Phe Met Ala Asp
80 85 90
Asp Ile Phe Ser Val Asn Ile Thr Asp Gln Ser Gly Asn Tyr Ser Gln
95 100 105
Glu Cys Gly Ser Phe Leu Leu Ala Glu Ser Ile Lys Pro Ala Pro Pro
110 115 120
Phe Asn Val Thr Val Thr Phe Ser Gly Gln Tyr Asn Ile Ser Trp Arg
125 130 135
Ser Asp Tyr Glu Asp Pro Ala Phe Tyr Met Leu Lys Gly Lys Leu Gln
140 145 150 155
Tyr Glu Leu Gln Tyr Arg Asn Arg Gly Asp Pro Trp Ala Val Ser Pro
160 165 170
Arg Arg Lys Leu Ile Ser Val Asp Ser Arg Ser Val Ser Leu Leu Pro

```
                175                 180                 185
Leu Glu Phe Arg Lys Asp Ser Ser Tyr Glu Leu Gln Val Arg Ala Gly
        190                 195                 200
Pro Met Pro Gly Ser Ser Tyr Gln Gly Thr Trp Ser Glu Trp Ser Asp
    205                 210                 215
Pro Val Ile Phe Gln Thr Gln Ser Glu Glu Leu Lys Glu Gly Trp Asn
220                 225                 230                 235
Pro His Leu Leu Leu Leu Leu Leu Leu Val Ile Val Phe Ile Pro Ala
                240                 245                 250
Phe Trp Ser Leu Lys Thr His Pro Leu Trp Arg Leu Trp Lys Lys Ile
            255                 260                 265
Trp Ala Val Pro Ser Pro Glu Arg Phe Phe Met Pro Leu Tyr Lys Gly
        270                 275                 280
Cys Ser Gly Asp Phe Lys Lys Trp Val Gly Ala Pro Phe Thr Gly Ser
    285                 290                 295
Ser Leu Glu Leu Gly Pro Trp Ser Pro Glu Val Pro Ser Thr Leu Glu
300                 305                 310                 315
Val Tyr Ser Cys His Pro Pro Ser Ser Pro Val Glu Cys Asp Phe Thr
                320                 325                 330
Ser Pro Gly Asp Glu Gly Pro Pro Arg Ser Tyr Leu Arg Gln Trp Val
            335                 340                 345
Val Ile Pro Pro Pro Leu Ser Ser Pro Gly Pro Gln Ala Ser
        350                 355                 360
```

[0107]

SEQ ID NO: 2

SEQUENCE LENGTH: 1884

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE TYPE: cDNA

SEQUENCE DESCRIPTION

GGCAGCCAGC GGCCTCAGAC AGACCCACTG GCGTCTCTCT GCTGAGTGAC CGTAAGCTCG 60

```
GCGTCTGGCC CTCTGCCTGC CTCTCCCTGA GTGTGGCTGA CAGCCACGCA GCTGTGTCTG      120
TCTGTCTGCG GCCCGTGCAT CCCTGCTGCG GCCGCCTGGT ACCTTCCTTG CCGTCTCTTT      180
CCTCTGTCTG CTGCTCTGTG GGACACCTGC CTGGAGGCCC AGCTGCCCGT CATCAGAGTG      240
ACAGGTCTTA TGACAGCCTG ATTGGTGACT CGGGCTGGGT GTGGATTCTC ACCCCAGGCC      300
TCTGCCTGCT TTCTCAGACC CTCATCTGTC ACCCCCACGC TGAACCCAGC TGCCACCCCC      360
AGAAGCCCAT CAGACTGCCC CCAGCACACG GAATGGATTT CTGAGAAAGA AGCCGAAACA      420
GAAGGCCCGT GGGAGTCAGC ATG CCG CGT GGC TGG GCC GCC CCC TTG CTC CTG      473
                     Met Pro Arg Gly Trp Ala Ala Pro Leu Leu Leu
                       1               5                  10
CTG CTG CTC CAG GGA GGC TGG GGC TGC CCC GAC CTC GTC TGC TAC ACC        521
Leu Leu Leu Gln Gly Gly Trp Gly Cys Pro Asp Leu Val Cys Tyr Thr
            15                  20                  25
GAT TAC CTC CAG ACG GTC ATC TGC ATC CTG GAA ATG TGG AAC CTC CAC        569
Asp Tyr Leu Gln Thr Val Ile Cys Ile Leu Glu Met Trp Asn Leu His
        30                  35                  40
CCC AGC ACG CTC ACC CTT ACC TGG CAA GAC CAG TAT GAA GAG CTG AAG        617
Pro Ser Thr Leu Thr Leu Thr Trp Gln Asp Gln Tyr Glu Glu Leu Lys
    45                  50                  55
GAC GAG GCC ACC TCC TGC AGC CTC CAC AGG TCG GCC CAC AAT GCC ACG        665
Asp Glu Ala Thr Ser Cys Ser Leu His Arg Ser Ala His Asn Ala Thr
60                  65                  70                  75
CAT GCC ACC TAC ACC TGC CAC ATG GAT GTA TTC CAC TTC ATG GCC GAC        713
His Ala Thr Tyr Thr Cys His Met Asp Val Phe His Phe Met Ala Asp
                80                  85                  90
GAC ATT TTC AGT GTC AAC ATC ACA GAC CAG TCT GGC AAC TAC TCC CAG        761
Asp Ile Phe Ser Val Asn Ile Thr Asp Gln Ser Gly Asn Tyr Ser Gln
            95                  100                 105
GAG TGT GGC AGC TTT CTC CTG GCT GAG AGC ATC AAG CCG GCT CCC CCT        809
Glu Cys Gly Ser Phe Leu Leu Ala Glu Ser Ile Lys Pro Ala Pro Pro
        110                 115                 120
TTC AAC GTG ACT GTG ACC TTC TCA GGA CAG TAT AAT ATC TCC TGG CGC        857
Phe Asn Val Thr Val Thr Phe Ser Gly Gln Tyr Asn Ile Ser Trp Arg
    125                 130                 135
TCA GAT TAC GAA GAC CCT GCC TTC TAC ATG CTG AAG GGC AAG CTT CAG        905
Ser Asp Tyr Glu Asp Pro Ala Phe Tyr Met Leu Lys Gly Lys Leu Gln
140                 145                 150                 155
```

```
TAT GAG CTG CAG TAC AGG AAC CGG GGA GAC CCC TGG GCT GTG AGT CCG        953
Tyr Glu Leu Gln Tyr Arg Asn Arg Gly Asp Pro Trp Ala Val Ser Pro
                160                 165                 170
AGG AGA AAG CTG ATC TCA GTG GAC TCA AGA AGT GTC TCC CTC CTC CCC       1001
Arg Arg Lys Leu Ile Ser Val Asp Ser Arg Ser Val Ser Leu Leu Pro
            175                 180                 185
CTG GAG TTC CGC AAA GAC TCG AGC TAT GAG CTG CAG GTG CGG GCA GGG       1049
Leu Glu Phe Arg Lys Asp Ser Ser Tyr Glu Leu Gln Val Arg Ala Gly
        190                 195                 200
CCC ATG CCT GGC TCC TCC TAC CAG GGG ACC TGG AGT GAA TGG AGT GAC       1097
Pro Met Pro Gly Ser Ser Tyr Gln Gly Thr Trp Ser Glu Trp Ser Asp
    205                 210                 215
CCG GTC ATC TTT CAG ACC CAG TCA GAG GAG TTA AAG GAA GGC TGG AAC       1145
Pro Val Ile Phe Gln Thr Gln Ser Glu Glu Leu Lys Glu Gly Trp Asn
220                 225                 230                 235
CCT CAC CTG CTG CTT CTC CTC CTG CTT GTC ATA GTC TTC ATT CCT GCC       1193
Pro His Leu Leu Leu Leu Leu Leu Leu Val Ile Val Phe Ile Pro Ala
                240                 245                 250
TTC TGG AGC CTG AAG ACC CAT CCA TTG TGG AGG CTA TGG AAG AAG ATA       1241
Phe Trp Ser Leu Lys Thr His Pro Leu Trp Arg Leu Trp Lys Lys Ile
            255                 260                 265
TGG GCC GTC CCC AGC CCT GAG CGG TTC TTC ATG CCC CTG TAC AAG GGC       1289
Trp Ala Val Pro Ser Pro Glu Arg Phe Phe Met Pro Leu Tyr Lys Gly
        270                 275                 280
TGC AGC GGA GAC TTC AAG AAA TGG GTG GGT GCA CCC TTC ACT GGC TCC       1337
Cys Ser Gly Asp Phe Lys Lys Trp Val Gly Ala Pro Phe Thr Gly Ser
    285                 290                 295
AGC CTG GAG CTG GGA CCC TGG AGC CCA GAG GTG CCC TCC ACC CTG GAG       1385
Ser Leu Glu Leu Gly Pro Trp Ser Pro Glu Val Pro Ser Thr Leu Glu
300                 305                 310                 315
GTG TAC AGC TGC CAC CCA CCC AGC AGC CCT GTG GAG TGT GAC TTC ACC       1433
Val Tyr Ser Cys His Pro Pro Ser Ser Pro Val Glu Cys Asp Phe Thr
                320                 325                 330
AGC CCC GGG GAC GAA GGA CCC CCC CGG AGC TAC CTC CGC CAG TGG GTG       1481
Ser Pro Gly Asp Glu Gly Pro Pro Arg Ser Tyr Leu Arg Gln Trp Val
            335                 340                 345
```

GTC ATT CCT CCG CCA CTT TCG AGC CCT GGA CCC CAG GCC AGC TAA 1526
Val Ile Pro Pro Pro Leu Ser Ser Pro Gly Pro Gln Ala Ser
350 355 360

TGAGGCTGAC TGGATGTCCA GAGCTGGCCA GGCCACTGGG CCCTGAGCCA GAGACAAGGT 1586
CACCTGGGCT GTGATGTGAA GACACCTGCA GCCTTTGGTC TCCTGGATGG GCCTTTGAGC 1646
CTGATGTTTA CAGTGTCTGT GTGTGTGTGC ATATGTGTGT GTGTGCATAT GCATGTGTGT 1706
GTGTGTGTGT GTCTTAGGTG CGCAGTGGCA TGTCCACGTG TGTGTGATTG CACGTGCCTG 1766
TGGGCCTGGG ATAATGCCCA TGGTACTCCA TGCATTCACC TGCCCTGTGC ATGTCTGGAC 1826
TCACGGAGCT CACCCATGTG CACAAGTGTG CACAGTAAAC GTGTTTGTGG TCAACAGA 1884

[0108]
SEQ ID NO: 3
SEQUENCE LENGTH: 144
SEQUENCE TYPE: amino acid
TOPOLOGY: linear
SEQUENCE TYPE: protein
SEQUENCE DESCRIPTION

Met Pro Arg Gly Trp Ala Ala Pro Leu Leu Leu
1 5 10
Leu Leu Leu Gln Gly Gly Trp Gly Cys Pro Asp Leu Val Cys Tyr Thr
15 20 25
Asp Tyr Leu Gln Thr Val Ile Cys Ile Leu Glu Met Trp Asn Leu His
30 35 40
Pro Ser Thr Leu Thr Leu Thr Trp Gln Asp Gln Tyr Glu Glu Leu Lys
45 50 55
Asp Glu Ala Thr Ser Cys Ser Leu His Arg Ser Ala His Asn Ala Thr
60 65 70 75
His Ala Thr Tyr Thr Cys His Met Asp Val Phe His Phe Met Ala Asp
80 85 90
Asp Ile Phe Ser Val Asn Ile Thr Asp Gln Ser Gly Asn Tyr Ser Gln
95 100 105
Glu Cys Gly Ser Phe Leu Leu Ala Glu Ser Lys Ser Glu Glu Lys Ala
110 115 120

Asp Leu Ser Gly Leu Lys Lys Cys Leu Pro Pro Pro Pro Gly Val Pro
125 130 135
Gln Arg Leu Glu Leu
140

[0109]

SEQ ID NO: 4
SEQUENCE LENGTH: 1729
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE TYPE: cDNA
SEQUENCE DESCRIPTION

```
GGCAGCCAGC GGCCTCAGAC AGACCCACTG GCGTCTCTCT GCTGAGTGAC CGTAAGCTCG      60
GCGTCTGGCC CTCTGCCTGC CTCTCCCTGA GTGTGGCTGA CAGCCACGCA GCTGTGTCTG     120
TCTGTCTGCG GCCCGTGCAT CCCTGCTGCG GCCGCCTGGT ACCTTCCTTG CCGTCTCTTT     180
CCTCTGTCTG CTGCTCTGTG GGACACCTGC CTGGAGGCCC AGCTGCCCGT CATCAGAGTG     240
ACAGGTCTTA TGACAGCCTG ATTGGTGACT CGGGCTGGGT GTGGATTCTC ACCCCAGGCC     300
TCTGCCTGCT TTCTCAGACC CTCATCTGTC ACCCCCACGC TGAACCCAGC TGCCACCCCC     360
AGAAGCCCAT CAGACTGCCC CCAGCACACG GAATGGATTT CTGAGAAAGA AGCCGAAACA     420
GAAGGCCCGT GGGAGTCAGC ATG CCG CGT GGC TGG GCC GCC CCC TTG CTC CTG      473
                     Met Pro Arg Gly Trp Ala Ala Pro Leu Leu Leu
                       1               5                  10
CTG CTG CTC CAG GGA GGC TGG GGC TGC CCC GAC CTC GTC TGC TAC ACC       521
Leu Leu Leu Gln Gly Gly Trp Gly Cys Pro Asp Leu Val Cys Tyr Thr
            15                  20                  25
GAT TAC CTC CAG ACG GTC ATC TGC ATC CTG GAA ATG TGG AAC CTC CAC       569
Asp Tyr Leu Gln Thr Val Ile Cys Ile Leu Glu Met Trp Asn Leu His
        30                  35                  40
CCC AGC ACG CTC ACC CTT ACC TGG CAA GAC CAG TAT GAA GAG CTG AAG       617
Pro Ser Thr Leu Thr Leu Thr Trp Gln Asp Gln Tyr Glu Glu Leu Lys
    45                  50                  55
GAC GAG GCC ACC TCC TGC AGC CTC CAC AGG TCG GCC CAC AAT GCC ACG       665
Asp Glu Ala Thr Ser Cys Ser Leu His Arg Ser Ala His Asn Ala Thr
```

```
 60                  65                  70                  75
CAT GCC ACC TAC ACC TGC CAC ATG GAT GTA TTC CAC TTC ATG GCC GAC        713
His Ala Thr Tyr Thr Cys His Met Asp Val Phe His Phe Met Ala Asp
                 80                  85                  90
GAC ATT TTC AGT GTC AAC ATC ACA GAC CAG TCT GGC AAC TAC TCC CAG        761
Asp Ile Phe Ser Val Asn Ile Thr Asp Gln Ser Gly Asn Tyr Ser Gln
            95                  100                 105
GAG TGT GGC AGC TTT CTC CTG GCT GAG AGC AAG TCC GAG GAG AAA GCT        809
Glu Cys Gly Ser Phe Leu Leu Ala Glu Ser Lys Ser Glu Glu Lys Ala
        110                 115                 120
GAT CTC AGT GGA CTC AAG AAG TGT CTC CCT CCT CCC CCT GGA GTT CCG        857
Asp Leu Ser Gly Leu Lys Lys Cys Leu Pro Pro Pro Pro Gly Val Pro
    125                 130                 135
CAA AGA CTC GAG CTA TGAGCTGCAG GTGCGGGCAG GGCCCATGCC TGGCTCCTCC        912
Gln Arg Leu Glu Leu
140
TACCAGGGGA CCTGGAGTGA ATGGAGTGAC CCGGTCATCT TTCAGACCCA GTCAGAGGAG      972
TTAAAGGAAG GCTGGAACCC TCACCTGCTG CTTCTCCTCC TGCTTGTCAT AGTCTTCATT     1032
CCTGCCTTCT GGAGCCTGAA GACCCATCCA TTGTGGAGGC TATGGAAGAA GATATGGGCC     1092
GTCCCCAGCC CTGAGCGGTT CTTCATGCCC CTGTACAAGG GCTGCAGCGG AGACTTCAAG     1152
AAATGGGTGG GTGCACCCTT CACTGGCTCC AGCCTGGAGC TGGGACCCTG GAGCCCAGAG     1212
GTGCCCTCCA CCCTGGAGGT GTACAGCTGC CACCCACCCA GCAGCCCTGT GGAGTGTGAC     1272
TTCACCAGCC CCGGGGACGA AGGACCCCCC CGGAGCTACC TCCGCCAGTG GGTGGTCATT     1332
CCTCCGCCAC TTTCGAGCCC TGGACCCCAG GCCAGCTAAT GAGGCTGACT GGATGTCCAG     1392
AGCTGGCCAG GCCACTGGGC CCTGAGCCAG AGACAAGGTC ACCTGGGCTG TGATGTGAAG     1452
ACACCTGCAG CCTTTGGTCT CCTGGATGGG CCTTTGAGCC TGATGTTTAC AGTGTCTGTG     1512
TGTGTGTGCA TATGTGTGTG TGTGCATATG CATGTGTGTG TGTGTGTGTG TCTTAGGTGC     1572
GCAGTGGCAT GTCCACGTGT GTGTGATTGC ACGTGCCTGT GGGCCTGGGA TAATGCCCAT     1632
GGTACTCCAT GCATTCACCT GCCCTGTGCA TGTCTGGACT CACGGAGCTC ACCCATGTGC     1692
ACAAGTGTGC ACAGTAAACG TGTTTGTGGT CAACAGA                              1729
```

[0110]

SEQ ID NO: 5

SEQUENCE LENGTH: 237

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

SEQUENCE TYPE: protein

SEQUENCE DESCRIPTION

Met Pro Arg Met Pro Pro Thr Pro Ala Thr Trp Met Tyr Ser Thr Ser
1 5 10 15
Trp Pro Thr Thr Phe Ser Val Ser Thr Ser Gln Thr Ser Leu Ala Thr
20 25 30
Thr Pro Arg Ser Val Ala Ala Phe Ser Trp Leu Arg Ala Ser Pro Arg
35 40 45
Arg Lys Leu Ile Ser Val Asp Ser Arg Ser Val Ser Leu Leu Pro Leu
50 55 60
Glu Phe Arg Lys Asp Ser Ser Tyr Glu Leu Gln Val Arg Ala Gly Pro
65 70 75 80
Met Pro Gly Ser Ser Tyr Gln Gly Thr Trp Ser Glu Trp Ser Asp Pro
85 90 95
Val Ile Phe Gln Thr Gln Ser Glu Glu Leu Lys Glu Gly Trp Asn Pro
100 105 110
His Leu Leu Leu Leu Leu Leu Leu Val Ile Val Phe Ile Pro Ala Phe
115 120 125
Trp Ser Leu Lys Thr His Pro Leu Trp Arg Leu Trp Lys Lys Ile Trp
130 135 140
Ala Val Pro Ser Pro Glu Arg Phe Phe Met Pro Leu Tyr Lys Gly Cys
145 150 155 160
Ser Gly Asp Phe Lys Lys Trp Val Gly Ala Pro Phe Thr Gly Ser Ser
165 170 175
Leu Glu Leu Gly Pro Trp Ser Pro Glu Val Pro Ser Thr Leu Glu Val
180 185 190
Tyr Ser Cys His Pro Pro Ser Ser Pro Val Glu Cys Asp Phe Thr Ser
195 200 205
Pro Gly Asp Glu Gly Pro Pro Arg Ser Tyr Leu Arg Gln Trp Val Val
210 215 220
Ile Pro Pro Pro Leu Ser Ser Pro Gly Pro Gln Ala Ser
225 230 235

[0111]

SEQ ID NO: 6

SEQUENCE LENGTH: 1729

SEQUENCE TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE TYPE: cDNA

SEQUENCE DESCRIPTION

```
GGCAGCCAGC GGCCTCAGAC AGACCCACTG GCGTCTCTCT GCTGAGTGAC CGTAAGCTCG      60
GCGTCTGGCC CTCTGCCTGC CTCTCCCTGA GTGTGGCTGA CAGCCACGCA GCTGTGTCTG     120
TCTGTCTGCG GCCCGTGCAT CCCTGCTGCG GCCGCCTGGT ACCTTCCTTG CCGTCTCTTT     180
CCTCTGTCTG CTGCTCTGTG GGACACCTGC CTGGAGGCCC AGCTGCCCGT CATCAGAGTG     240
ACAGGTCTTA TGACAGCCTG ATTGGTGACT CGGGCTGGGT GTGGATTCTC ACCCCAGGCC     300
TCTGCCTGCT TTCTCAGACC CTCATCTGTC ACCCCCACGC TGAACCCAGC TGCCACCCCC     360
AGAAGCCCAT CAGACTGCCC CCAGCACACG GAATGGATTT CTGAGAAAGA AGCCGAAACA     420
GAAGGCCCGT GGGAGTCAGC ATGCCGCGTG GCTGGGCCGC CCCCTTGCTC CTGCTGCTGC     480
TCCAGGGAGG CTGGGGCTGC CCCGACCTCG TCTGCTACAC CGATTACCTC CAGACGGTCA     540
TCTGCATCCT GGAAATGTGG AACCTCCACC CCAGCACGCT CACCCTTACC TGGCAAGACC     600
AGTATGAAGA GCTGAAGGAC GAGGCCACCT CCTGCAGCCT CCACAGGTCG GCCCACAA       658
ATG CCA CGC ATG CCA CCT ACA CCT GCC ACA TGG ATG TAT TCC ACT TCA       705
Met Pro Arg Met Pro Pro Thr Pro Ala Thr Trp Met Tyr Ser Thr Ser
 1               5                  10                  15
TGG CCG ACG ACA TTT TCA GTG TCA ACA TCA CAG ACC AGT CTG GCA ACT       753
Trp Pro Thr Thr Phe Ser Val Ser Thr Ser Gln Thr Ser Leu Ala Thr
            20                  25                  30
ACT CCC AGG AGT GTG GCA GCT TTC TCC TGG CTG AGA GCA AGT CCG AGG       801
Thr Pro Arg Ser Val Ala Ala Phe Ser Trp Leu Arg Ala Ser Pro Arg
        35                  40                  45
AGA AAG CTG ATC TCA GTG GAC TCA AGA AGT GTC TCC CTC CTC CCC CTG       849
Arg Lys Leu Ile Ser Val Asp Ser Arg Ser Val Ser Leu Leu Pro Leu
    50                  55                  60
GAG TTC CGC AAA GAC TCG AGC TAT GAG CTG CAG GTG CGG GCA GGG CCC       897
```

```
Glu Phe Arg Lys Asp Ser Ser Tyr Glu Leu Gln Val Arg Ala Gly Pro
 65                  70                  75                  80
ATG CCT GGC TCC TCC TAC CAG GGG ACC TGG AGT GAA TGG AGT GAC CCG        945
Met Pro Gly Ser Ser Tyr Gln Gly Thr Trp Ser Glu Trp Ser Asp Pro
                 85                  90                  95
GTC ATC TTT CAG ACC CAG TCA GAG GAG TTA AAG GAA GGC TGG AAC CCT        993
Val Ile Phe Gln Thr Gln Ser Glu Glu Leu Lys Glu Gly Trp Asn Pro
            100                 105                 110
CAC CTG CTG CTT CTC CTC CTG CTT GTC ATA GTC TTC ATT CCT GCC TTC       1041
His Leu Leu Leu Leu Leu Leu Leu Val Ile Val Phe Ile Pro Ala Phe
        115                 120                 125
TGG AGC CTG AAG ACC CAT CCA TTG TGG AGG CTA TGG AAG AAG ATA TGG       1089
Trp Ser Leu Lys Thr His Pro Leu Trp Arg Leu Trp Lys Lys Ile Trp
    130                 135                 140
GCC GTC CCC AGC CCT GAG CGG TTC TTC ATG CCC CTG TAC AAG GGC TGC       1137
Ala Val Pro Ser Pro Glu Arg Phe Phe Met Pro Leu Tyr Lys Gly Cys
145                 150                 155                 160
AGC GGA GAC TTC AAG AAA TGG GTG GGT GCA CCC TTC ACT GGC TCC AGC       1185
Ser Gly Asp Phe Lys Lys Trp Val Gly Ala Pro Phe Thr Gly Ser Ser
                165                 170                 175
CTG GAG CTG GGA CCC TGG AGC CCA GAG GTG CCC TCC ACC CTG GAG GTG       1233
Leu Glu Leu Gly Pro Trp Ser Pro Glu Val Pro Ser Thr Leu Glu Val
            180                 185                 190
TAC AGC TGC CAC CCA CCC AGC AGC CCT GTG GAG TGT GAC TTC ACC AGC       1281
Tyr Ser Cys His Pro Pro Ser Ser Pro Val Glu Cys Asp Phe Thr Ser
        195                 200                 205
CCC GGG GAC GAA GGA CCC CCC CGG AGC TAC CTC CGC CAG TGG GTG GTC       1329
Pro Gly Asp Glu Gly Pro Pro Arg Ser Tyr Leu Arg Gln Trp Val Val
    210                 215                 220
ATT CCT CCG CCA CTT TCG AGC CCT GGA CCC CAG GCC AGC TAATGAGGCT        1378
Ile Pro Pro Pro Leu Ser Ser Pro Gly Pro Gln Ala Ser
225                 230                 235
GACTGGATGT CCAGAGCTGG CCAGGCCACT GGGCCCTGAG CCAGAGACAA GGTCACCTGG     1438
GCTGTGATGT GAAGACACCT GCAGCCTTTG GTCTCCTGGA TGGGCCTTTG AGCCTGATGT     1498
TTACAGTGTC TGTGTGTGTG TGCATATGTG TGTGTGTGCA TATGCATGTG TGTGTGTGTG     1558
TGTGTCTTAG GTGCGCAGTG GCATGTCCAC GTGTGTGTGA TTGCACGTGC CTGTGGGCCT     1618
```

GGGATAATGC CCATGGTACT CCATGCATTC ACCTGCCCTG TGCATGTCTG GACTCACGGA 1678

GCTCACCCAT GTGCACAAGT GTGCACAGTA AACGTGTTTG TGGTCAACAGA 1729

[0112]

SEQ ID NO: 7

SEQUENCE LENGTH: 538

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

SEQUENCE TYPE: protein

SEQUENCE DESCRIPTION

Met Pro Arg Gly Trp Ala Ala Pro Leu Leu Leu
1 5 10

Leu Leu Leu Gln Gly Gly Trp Gly Cys Pro Asp Leu Val Cys Tyr Thr
15 20 25

Asp Tyr Leu Gln Thr Val Ile Cys Ile Leu Glu Met Trp Asn Leu His
30 35 40

Pro Ser Thr Leu Thr Leu Thr Trp Gln Asp Gln Tyr Glu Glu Leu Lys
45 50 55

Asp Glu Ala Thr Ser Cys Ser Leu His Arg Ser Ala His Asn Ala Thr
60 65 70 75

His Ala Thr Tyr Thr Cys His Met Asp Val Phe His Phe Met Ala Asp
80 85 90

Asp Ile Phe Ser Val Asn Ile Thr Asp Gln Ser Gly Asn Tyr Ser Gln
95 100 105

Glu Cys Gly Ser Phe Leu Leu Ala Glu Ser Ile Lys Pro Ala Pro Pro
110 115 120

Phe Asn Val Thr Val Thr Phe Ser Gly Gln Tyr Asn Ile Ser Trp Arg
125 130 135

Ser Asp Tyr Glu Asp Pro Ala Phe Tyr Met Leu Lys Gly Lys Leu Gln
140 145 150 155

Tyr Glu Leu Gln Tyr Arg Asn Arg Gly Asp Pro Trp Ala Val Ser Pro
160 165 170

Arg Arg Lys Leu Ile Ser Val Asp Ser Arg Ser Val Ser Leu Leu Pro
175 180 185

Leu Glu Phe Arg Lys Asp Ser Ser Tyr Glu Leu Gln Val Arg Ala Gly
190 195 200
Pro Met Pro Gly Ser Ser Tyr Gln Gly Thr Trp Ser Glu Trp Ser Asp
205 210 215
Pro Val Ile Phe Gln Thr Gln Ser Glu Glu Leu Lys Glu Gly Trp Asn
220 225 230 235
Pro His Leu Leu Leu Leu Leu Leu Leu Val Ile Val Phe Ile Pro Ala
240 245 250
Phe Trp Ser Leu Lys Thr His Pro Leu Trp Arg Leu Trp Lys Lys Ile
255 260 265
Trp Ala Val Pro Ser Pro Glu Arg Phe Phe Met Pro Leu Tyr Lys Gly
270 275 280
Cys Ser Gly Asp Phe Lys Lys Trp Val Gly Ala Pro Phe Thr Gly Ser
285 290 295
Ser Leu Glu Leu Gly Pro Trp Ser Pro Glu Val Pro Ser Thr Leu Glu
300 305 310 315
Val Tyr Ser Cys His Pro Pro Arg Ser Pro Ala Lys Arg Leu Gln Leu
320 325 330
Thr Glu Leu Gln Glu Pro Ala Glu Leu Val Glu Ser Asp Gly Val Pro
335 340 345
Lys Pro Ser Phe Trp Pro Thr Ala Gln Asn Ser Gly Gly Ser Ala Tyr
350 355 360
Ser Glu Glu Arg Asp Arg Pro Tyr Gly Leu Val Ser Ile Asp Thr Val
365 370 375
Thr Val Leu Asp Ala Glu Gly Pro Cys Thr Trp Pro Cys Ser Cys Glu
380 385 390 395
Asp Asp Gly Tyr Pro Ala Leu Asp Leu Asp Ala Gly Leu Glu Pro Ser
400 405 410
Pro Gly Leu Glu Asp Pro Leu Leu Asp Ala Gly Thr Thr Val Leu Ser
415 420 425
Cys Gly Cys Val Ser Ala Gly Ser Pro Gly Leu Gly Gly Pro Leu Gly
430 435 440
Ser Leu Leu Asp Arg Leu Lys Pro Pro Leu Ala Asp Gly Glu Asp Trp
445 450 455
Ala Gly Gly Leu Pro Trp Gly Gly Arg Ser Pro Gly Gly Val Ser Glu
460 465 470 475

Ser Glu Ala Gly Ser Pro Leu Ala Gly Leu Asp Met Asp Thr Phe Asp
480 485 490
Ser Gly Phe Val Gly Ser Asp Cys Ser Ser Pro Val Glu Cys Asp Phe
495 500 505
Thr Ser Pro Gly Asp Glu Gly Pro Pro Arg Ser Tyr Leu Arg Gln Trp
510 515 520
Val Val Ile Pro Pro Pro Leu Ser Ser Pro Gly Pro Gln Ala Ser
525 530 535

[0113]

SEQ ID NO: 8
SEQUENCE LENGTH: 2415
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE TYPE: cDNA
SEQUENCE DESCRIPTION

```
GGCAGCCAGC GGCCTCAGAC AGACCCACTG GCGTCTCTCT GCTGAGTGAC CGTAAGCTCG      60
GCGTCTGGCC CTCTGCCTGC CTCTCCCTGA GTGTGGCTGA CAGCCACGCA GCTGTGTCTG     120
TCTGTCTGCG GCCCGTGCAT CCCTGCTGCG GCCGCCTGGT ACCTTCCTTG CCGTCTCTTT     180
CCTCTGTCTG CTGCTCTGTG GGACACCTGC CTGGAGGCCC AGCTGCCCGT CATCAGAGTG     240
ACAGGTCTTA TGACAGCCTG ATTGGTGACT CGGGCTGGGT GTGGATTCTC ACCCCAGGCC     300
TCTGCCTGCT TTCTCAGACC CTCATCTGTC ACCCCCACGC TGAACCCAGC TGCCACCCCC     360
AGAAGCCCAT CAGACTGCCC CCAGCACACG GAATGGATTT CTGAGAAAGA AGCCGAAACA     420
GAAGGCCCGT GGGAGTCAGC ATG CCG CGT GGC TGG GCC GCC CCC TTG CTC CTG      473
                     Met Pro Arg Gly Trp Ala Ala Pro Leu Leu Leu
                       1               5                  10
CTG CTG CTC CAG GGA GGC TGG GGC TGC CCC GAC CTC GTC TGC TAC ACC       521
Leu Leu Leu Gln Gly Gly Trp Gly Cys Pro Asp Leu Val Cys Tyr Thr
            15                  20                  25
GAT TAC CTC CAG ACG GTC ATC TGC ATC CTG GAA ATG TGG AAC CTC CAC       569
Asp Tyr Leu Gln Thr Val Ile Cys Ile Leu Glu Met Trp Asn Leu His
```

```
          30                  35                  40
CCC AGC ACG CTC ACC CTT ACC TGG CAA GAC CAG TAT GAA GAG CTG AAG        617
Pro Ser Thr Leu Thr Leu Thr Trp Gln Asp Gln Tyr Glu Glu Leu Lys
     45                  50                  55
GAC GAG GCC ACC TCC TGC AGC CTC CAC AGG TCG GCC CAC AAT GCC ACG        665
Asp Glu Ala Thr Ser Cys Ser Leu His Arg Ser Ala His Asn Ala Thr
 60                  65                  70                  75
CAT GCC ACC TAC ACC TGC CAC ATG GAT GTA TTC CAC TTC ATG GCC GAC        713
His Ala Thr Tyr Thr Cys His Met Asp Val Phe His Phe Met Ala Asp
                 80                  85                  90
GAC ATT TTC AGT GTC AAC ATC ACA GAC CAG TCT GGC AAC TAC TCC CAG        761
Asp Ile Phe Ser Val Asn Ile Thr Asp Gln Ser Gly Asn Tyr Ser Gln
             95                 100                 105
GAG TGT GGC AGC TTT CTC CTG GCT GAG AGC ATC AAG CCG GCT CCC CCT        809
Glu Cys Gly Ser Phe Leu Leu Ala Glu Ser Ile Lys Pro Ala Pro Pro
        110                 115                 120
TTC AAC GTG ACT GTG ACC TTC TCA GGA CAG TAT AAT ATC TCC TGG CGC        857
Phe Asn Val Thr Val Thr Phe Ser Gly Gln Tyr Asn Ile Ser Trp Arg
    125                 130                 135
TCA GAT TAC GAA GAC CCT GCC TTC TAC ATG CTG AAG GGC AAG CTT CAG        905
Ser Asp Tyr Glu Asp Pro Ala Phe Tyr Met Leu Lys Gly Lys Leu Gln
140                 145                 150                 155
TAT GAG CTG CAG TAC AGG AAC CGG GGA GAC CCC TGG GCT GTG AGT CCG        953
Tyr Glu Leu Gln Tyr Arg Asn Arg Gly Asp Pro Trp Ala Val Ser Pro
                160                 165                 170
AGG AGA AAG CTG ATC TCA GTG GAC TCA AGA AGT GTC TCC CTC CTC CCC       1001
Arg Arg Lys Leu Ile Ser Val Asp Ser Arg Ser Val Ser Leu Leu Pro
            175                 180                 185
CTG GAG TTC CGC AAA GAC TCG AGC TAT GAG CTG CAG GTG CGG GCA GGG       1049
Leu Glu Phe Arg Lys Asp Ser Ser Tyr Glu Leu Gln Val Arg Ala Gly
        190                 195                 200
CCC ATG CCT GGC TCC TCC TAC CAG GGG ACC TGG AGT GAA TGG AGT GAC       1097
Pro Met Pro Gly Ser Ser Tyr Gln Gly Thr Trp Ser Glu Trp Ser Asp
    205                 210                 215
CCG GTC ATC TTT CAG ACC CAG TCA GAG GAG TTA AAG GAA GGC TGG AAC       1145
Pro Val Ile Phe Gln Thr Gln Ser Glu Glu Leu Lys Glu Gly Trp Asn
```

```
220                 225                 230                 235
CCT CAC CTG CTG CTT CTC CTC CTG CTT GTC ATA GTC TTC ATT CCT GCC      1193
Pro His Leu Leu Leu Leu Leu Leu Leu Val Ile Val Phe Ile Pro Ala
                240                 245                 250
TTC TGG AGC CTG AAG ACC CAT CCA TTG TGG AGG CTA TGG AAG AAG ATA      1241
Phe Trp Ser Leu Lys Thr His Pro Leu Trp Arg Leu Trp Lys Lys Ile
            255                 260                 265
TGG GCC GTC CCC AGC CCT GAG CGG TTC TTC ATG CCC CTG TAC AAG GGC      1289
Trp Ala Val Pro Ser Pro Glu Arg Phe Phe Met Pro Leu Tyr Lys Gly
        270                 275                 280
TGC AGC GGA GAC TTC AAG AAA TGG GTG GGT GCA CCC TTC ACT GGC TCC      1337
Cys Ser Gly Asp Phe Lys Lys Trp Val Gly Ala Pro Phe Thr Gly Ser
    285                 290                 295
AGC CTG GAG CTG GGA CCC TGG AGC CCA GAG GTG CCC TCC ACC CTG GAG      1385
Ser Leu Glu Leu Gly Pro Trp Ser Pro Glu Val Pro Ser Thr Leu Glu
300                 305                 310                 315
GTG TAC AGC TGC CAC CCA CCA CGG AGC CCG GCC AAG AGG CTG CAG CTC      1433
Val Tyr Ser Cys His Pro Pro Arg Ser Pro Ala Lys Arg Leu Gln Leu
                320                 325                 330
ACG GAG CTA CAA GAA CCA GCA GAG CTG GTG GAG TCT GAC GGT GTG CCC      1481
Thr Glu Leu Gln Glu Pro Ala Glu Leu Val Glu Ser Asp Gly Val Pro
            335                 340                 345
AAG CCC AGC TTC TGG CCG ACA GCC CAG AAC TCG GGG GGC TCA GCT TAC      1529
Lys Pro Ser Phe Trp Pro Thr Ala Gln Asn Ser Gly Gly Ser Ala Tyr
        350                 355                 360
AGT GAG GAG AGG GAT CGG CCA TAC GGC CTG GTG TCC ATT GAC ACA GTG      1577
Ser Glu Glu Arg Asp Arg Pro Tyr Gly Leu Val Ser Ile Asp Thr Val
    365                 370                 375
ACT GTG CTA GAT GCA GAG GGG CCA TGC ACC TGG CCC TGC AGC TGT GAG      1625
Thr Val Leu Asp Ala Glu Gly Pro Cys Thr Trp Pro Cys Ser Cys Glu
380                 385                 390                 395
GAT GAC GGC TAC CCA GCC CTG GAC CTG GAT GCT GGC CTG GAG CCC AGC      1673
Asp Asp Gly Tyr Pro Ala Leu Asp Leu Asp Ala Gly Leu Glu Pro Ser
                400                 405                 410
CCA GGC CTA GAG GAC CCA CTC TTG GAT GCA GGG ACC ACA GTC CTG TCC      1721
Pro Gly Leu Glu Asp Pro Leu Leu Asp Ala Gly Thr Thr Val Leu Ser
```

```
            415                 420                 425
TGT GGC TGT GTC TCA GCT GGC AGC CCT GGG CTA GGA GGG CCC CTG GGA      1769
Cys Gly Cys Val Ser Ala Gly Ser Pro Gly Leu Gly Gly Pro Leu Gly
        430                 435                 440
AGC CTC CTG GAC AGA CTA AAG CCA CCC CTT GCA GAT GGG GAG GAC TGG      1817
Ser Leu Leu Asp Arg Leu Lys Pro Pro Leu Ala Asp Gly Glu Asp Trp
    445                 450                 455
GCT GGG GGA CTG CCC TGG GGT GGC CGG TCA CCT GGA GGG GTC TCA GAG      1865
Ala Gly Gly Leu Pro Trp Gly Gly Arg Ser Pro Gly Gly Val Ser Glu
460                 465                 470                 475
AGT GAG GCG GGC TCA CCC CTG GCC GGC CTG GAT ATG GAC ACG TTT GAC      1913
Ser Glu Ala Gly Ser Pro Leu Ala Gly Leu Asp Met Asp Thr Phe Asp
                480                 485                 490
AGT GGC TTT GTG GGC TCT GAC TGC AGC AGC CCT GTG GAG TGT GAC TTC      1961
Ser Gly Phe Val Gly Ser Asp Cys Ser Ser Pro Val Glu Cys Asp Phe
            495                 500                 505
ACC AGC CCC GGG GAC GAA GGA CCC CCC CGG AGC TAC CTC CGC CAG TGG      2009
Thr Ser Pro Gly Asp Glu Gly Pro Pro Arg Ser Tyr Leu Arg Gln Trp
        510                 515                 520
GTG GTC ATT CCT CCG CCA CTT TCG AGC CCT GGA CCC CAG GCC AGC TAA      2057
Val Val Ile Pro Pro Pro Leu Ser Ser Pro Gly Pro Gln Ala Ser
    525                 530                 535
TGAGGCTGAC TGGATGTCCA GAGCTGGCCA GGCCACTGGG CCCTGAGCCA GAGACAAGGT    2117
CACCTGGGCT GTGATGTGAA GACACCTGCA GCCTTTGGTC TCCTGGATGG GCCTTTGAGC    2177
CTGATGTTTA CAGTGTCTGT GTGTGTGTGC ATATGTGTGT GTGTGCATAT GCATGTGTGT    2237
GTGTGTGTGT GTCTTAGGTG CGCAGTGGCA TGTCCACGTG TGTGTGATTG CACGTGCCTG    2297
TGGGCCTGGG ATAATGCCCA TGGTACTCCA TGCATTCACC TGCCCTGTGC ATGTCTGGAC    2357
TCACGGAGCT CACCCATGTG CACAAGTGTG CACAGTAAAC GTGTTTGTGG TCAACAGA      2415
```

[0114]

SEQ ID NO: 9

SEQUENCE LENGTH: 30
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE TYPE: other nucleic acids, synthetic DNA
SEQUENCE DESCRIPTION

CCGGCTCCCC CTTTCAACGT GACTGTGACC 30

[0115]
SEQ ID NO: 10
SEQUENCE LENGTH: 30
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE TYPE: other nucleic acids, synthetic DNA
SEQUENCE DESCRIPTION

GGCAAGCTTC AGTATGAGCT GCAGTACAGG 30

[0116]
SEQ ID NO: 11
SEQUENCE LENGTH: 30
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE TYPE: other nucleic acids, synthetic DNA
SEQUENCE DESCRIPTION

ACCCTCTGAC TGGGTCTGAA AGATGACCGG 30

[0117]
SEQ ID NO: 12
SEQUENCE LENGTH: 30
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE TYPE: other nucleic acids, synthetic DNA
SEQUENCE DESCRIPTION

CATGGGCCCT GCCCGCACCT GCAGCTCATA 30

[Brief Description of the Drawings]

[Fig. 1]

Fig. 1 is a schematic diagram showing the results of BlastX search where the query was 180 nucleotides of 40952-40966 including 40952-40966, the only probe sequence within the AC002303.

[Fig. 2]

Fig. 2 is a schematic diagram showing the results of BlastX scanning of 180 nucleotides in both the 5′ and 3′ directions, where the search centered on the 180 nucleotides of 40952-40966 containing 40952-40966, the only probe sequence within the AC002303.

[Fig. 3]

Fig. 3 shows the electrophoresis results of the amplification done by the RT-PCR method for the combinations of SN1/AS1, SN1/AS2, SN2/AS1, and SN2/AS2 primers using human fetal liver and skeletal muscle cDNA as templates.

[Fig. 4]

Fig. 4 shows the electrophoretic results of the 5′-RACE method and 3′-RACE method using human fetal liver cDNA as the template.

[Fig. 5]

Fig. 5 shows the nucleotide sequence and the amino acid sequence of NR8α cDNA.

[Fig. 6]

Fig. 6 shows the nucleotide sequence and the amino acid sequence of NR8β cDNA. Two possible open reading frames (ORF) are shown.

[Fig. 7]

Fig. 7 shows the nucleotide sequence and the amino acid sequence of NR8γ cDNA. The 177 amino acids inserted by alternative splicing are underlined.

[Fig. 8]

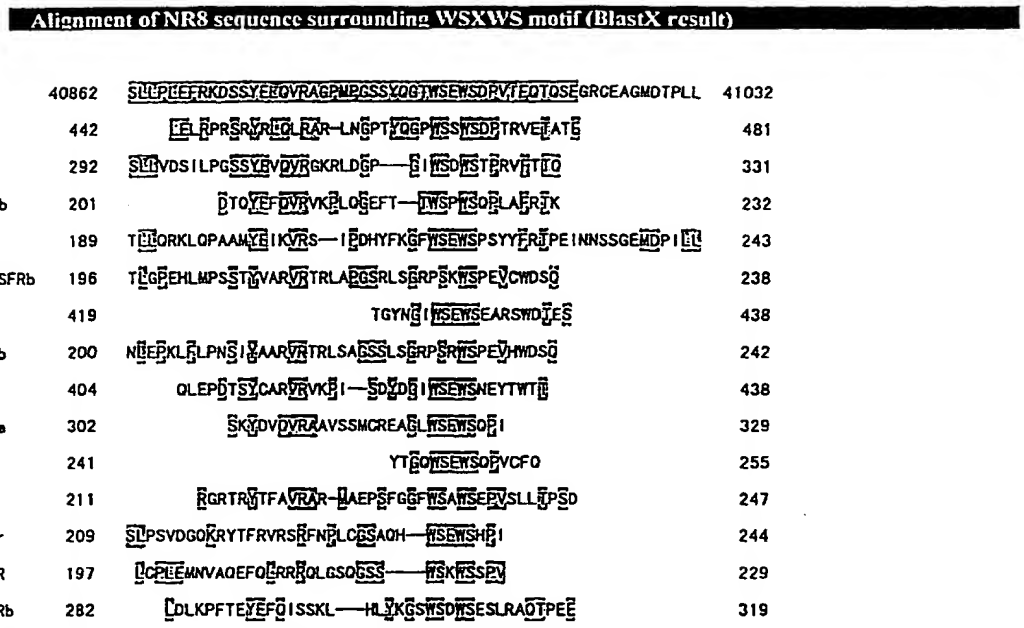
Fig. 8 shows the results of Northern blot analysis of NR8 expression in each organ.

[Fig. 9]

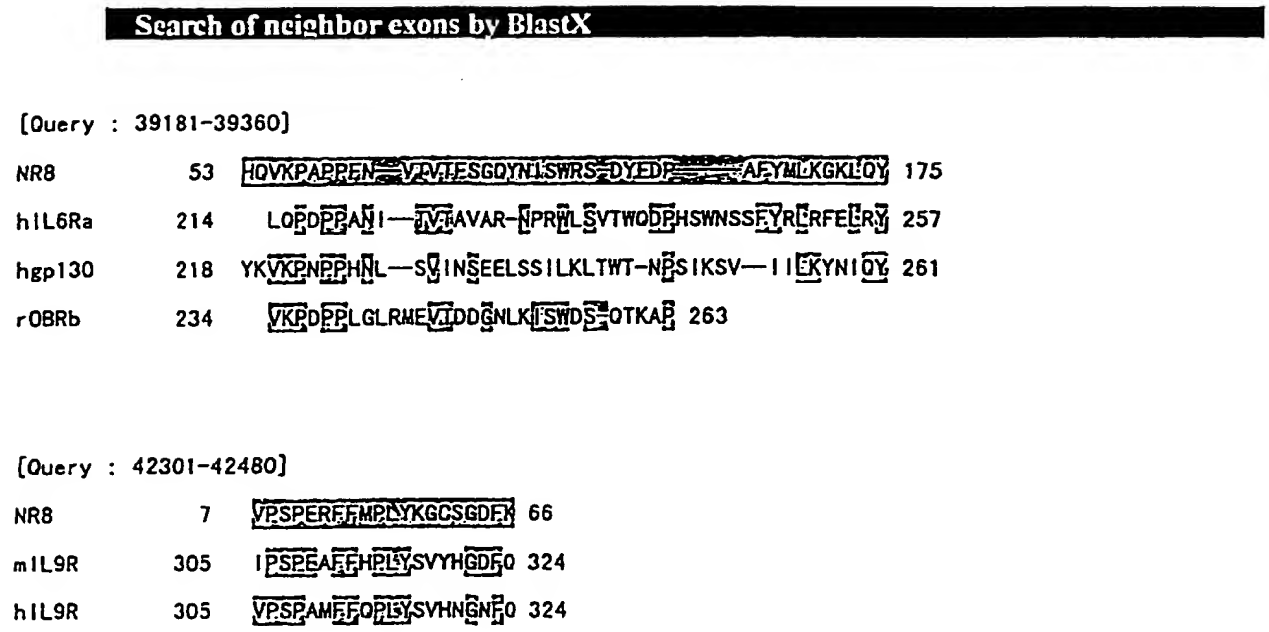
Fig. 9 is a schematic diagram showing the structure of the NR8 gene.

Drawings

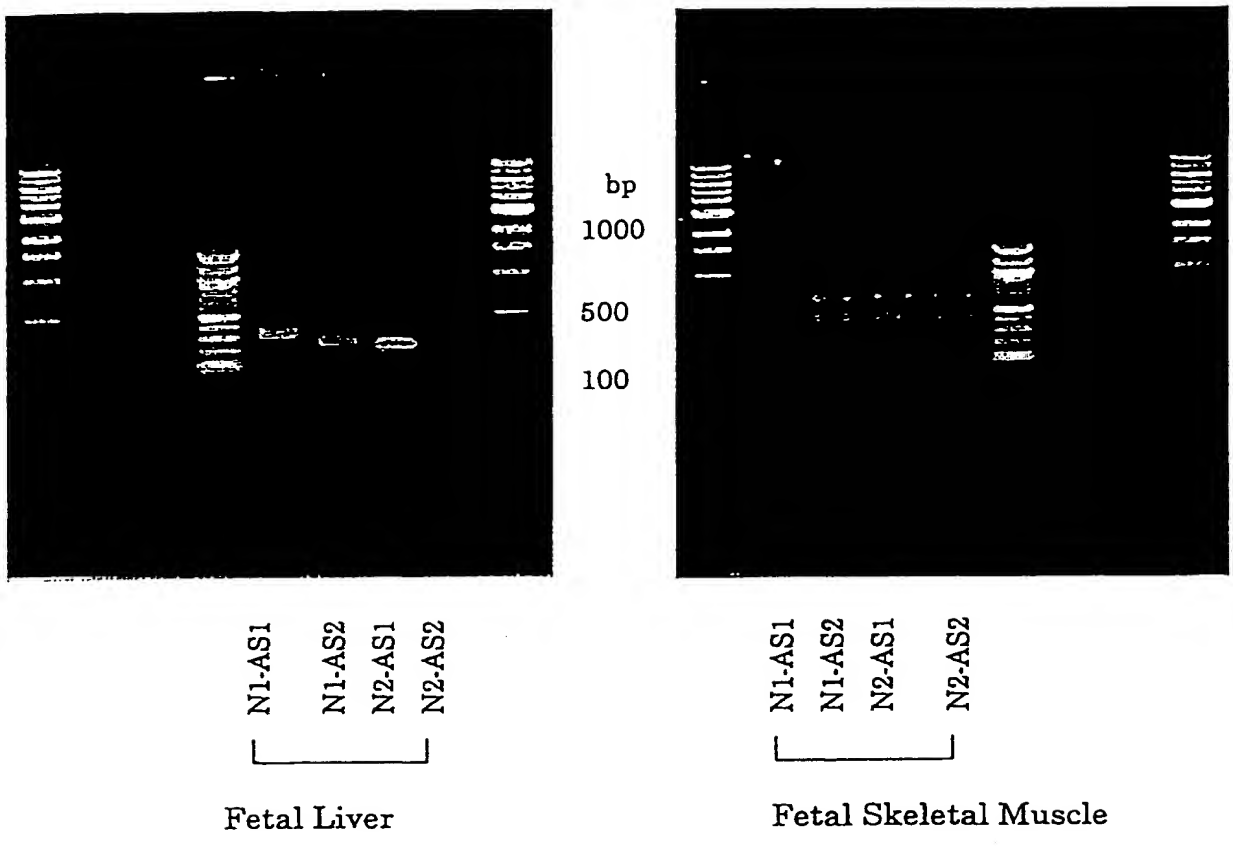
[Fig. 1]



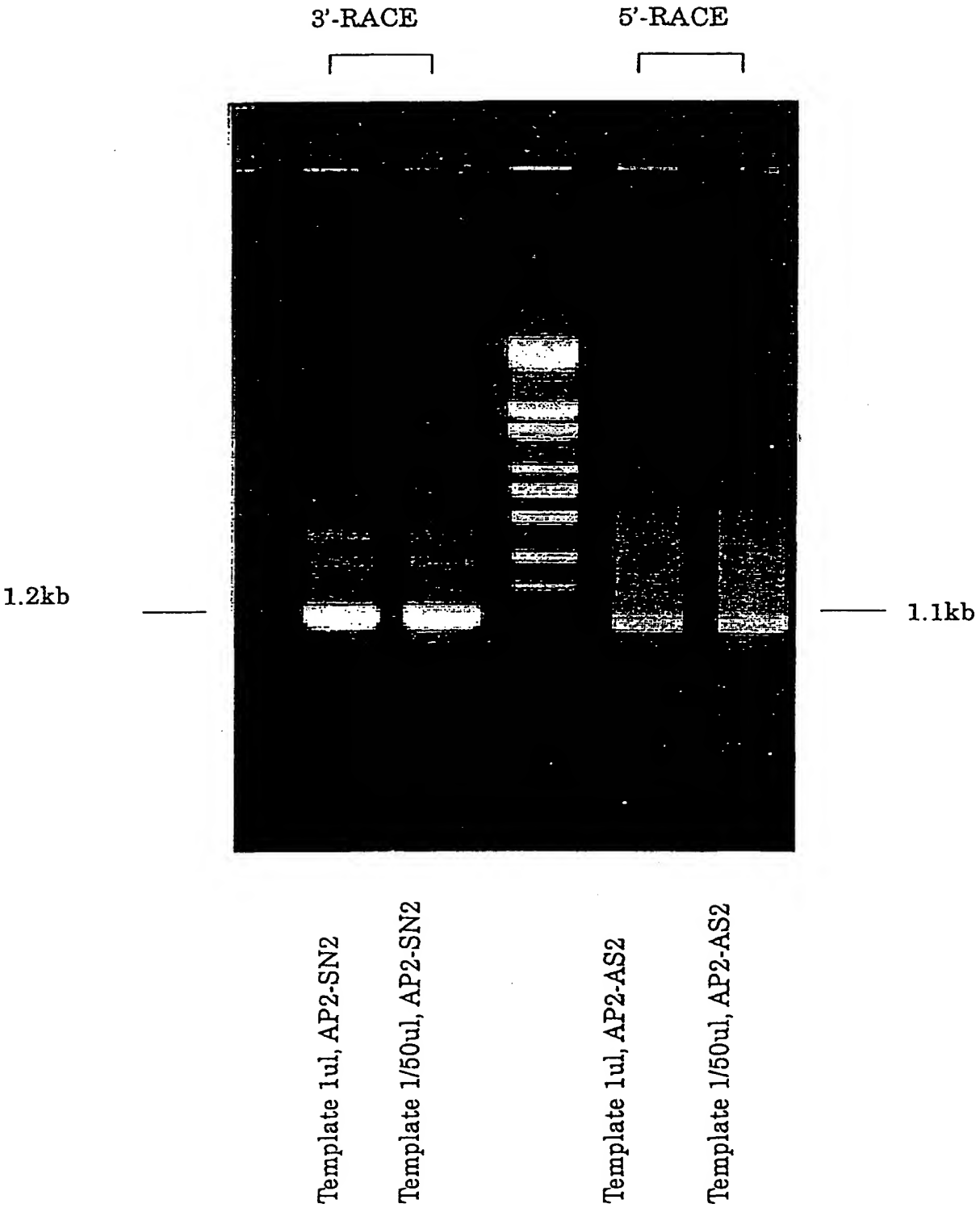
[Fig. 2]


Search of neighbor exons by BlastX
[Query : 39181-39360]
NR8 53 HQVKPAPPEN VDVTESGQYNLSWRSDYEDP AEYMLKGKLQY 175
hIL6Ra 214 LQPDPPANI IVTAVAR-NPRWLSVTWQDPHSWNSSFYRLRFELRY 257
hgp130 218 YKVKPNPPHNL SVINSEELSSILKLTWT-NPSIKSV IIEKYNIQY 261
rOBRb 234 VKPDPPLGLRMEVIDDGNLKISWDSQTKAP 263
[Query : 42301-42480]
NR8 7 VPSPERFFMPLYKGCSGDFK 66
mIL9R 305 IPSPEAFFHPLYSVYHGDFQ 324
hIL9R 305 VPSPAMFFQPLYSVHNGNFQ 324

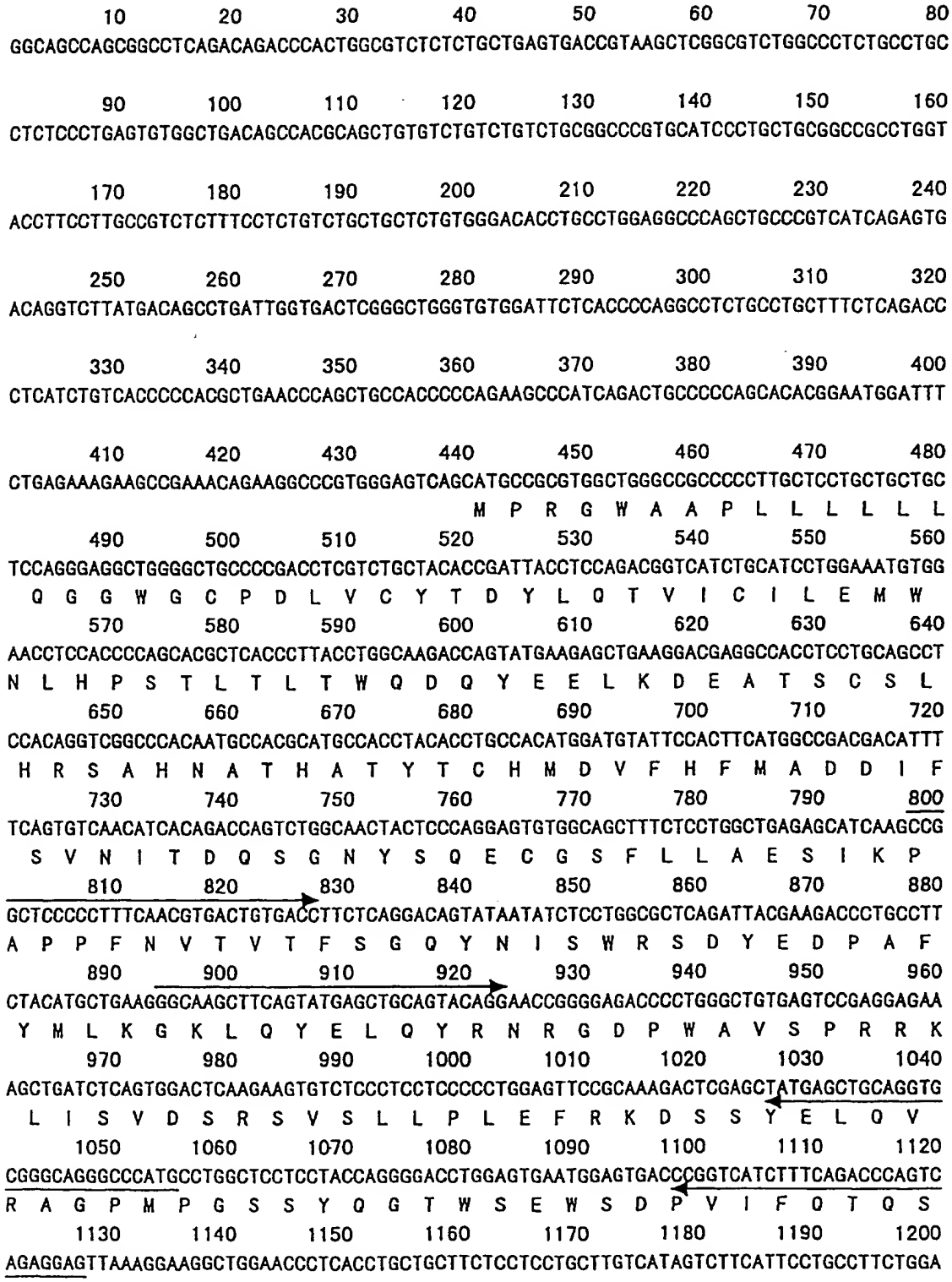
[Fig. 3]


bp
1000
500
100
N1-AS1
N1-AS2
N2-AS1
N2-AS2
Fetal Liver
N1-AS1
N1-AS2
N2-AS1
N2-AS2
Fetal Skeletal Muscle

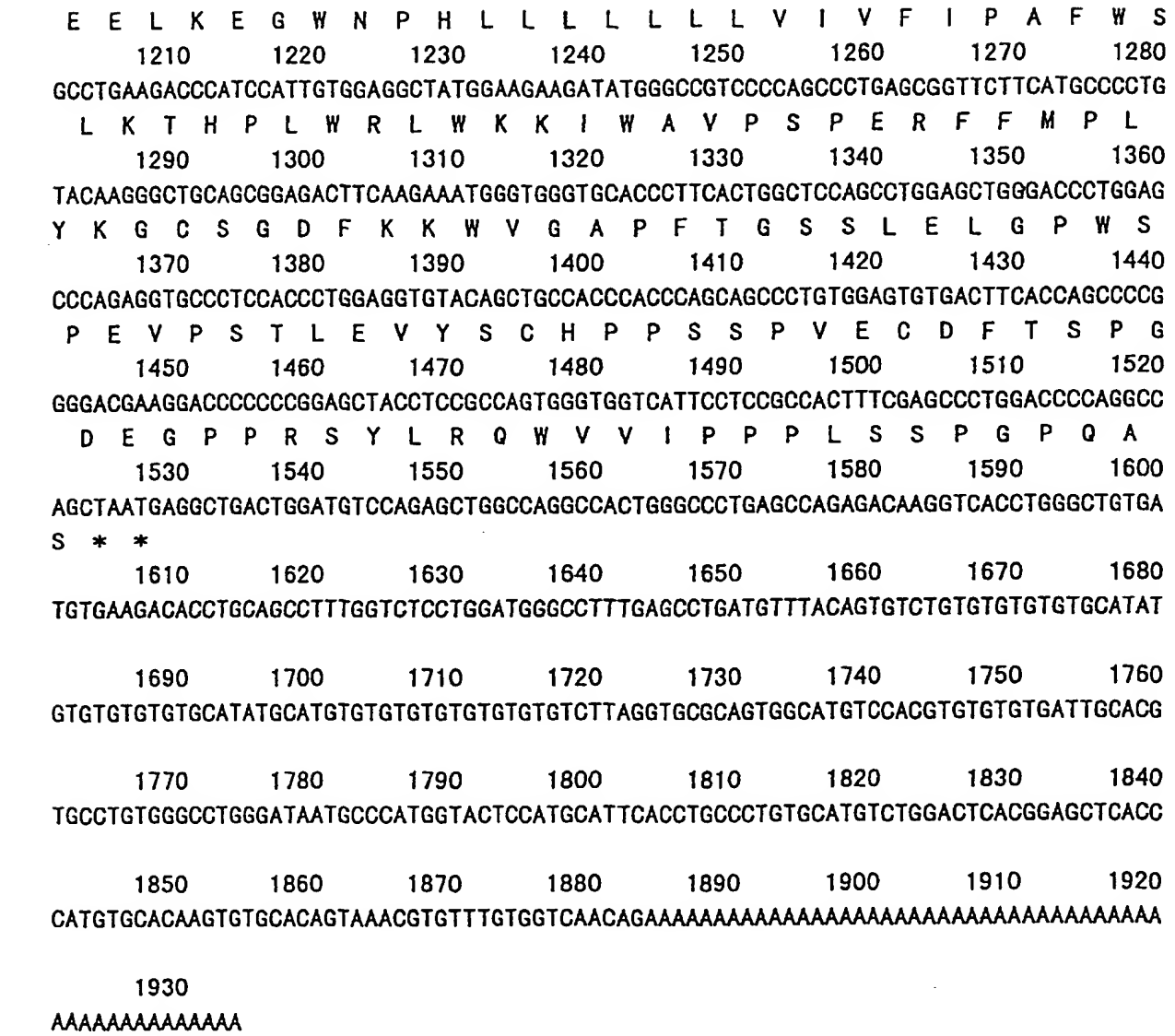
[Fig. 4]


3'-RACE
5'-RACE
1.2kb
1.1kb
Template 1ul, AP2-SN2
Template 1/50ul, AP2-SN2
Template 1ul, AP2-AS2
Template 1/50ul, AP2-AS2

[Fig. 5-1]


10 20 30 40 50 60 70 80
GGCAGCCAGCGGCCTCAGACAGACCCACTGGCGTCTCTCTGCTGAGTGACCGTAAGCTCGGCGTCTGGCCCTCTGCCTGC
90 100 110 120 130 140 150 160
CTCTCCCTGAGTGTGGCTGACAGCCACGCAGCTGTGTCTGTCTGTCTGCGGCCCGTGCATCCCTGCTGCGGCCGCCTGGT
170 180 190 200 210 220 230 240
ACCTTCCTTGCCGTCTCTTTCCTCTGTCTGCTGCTCTGTGGGACACCTGCCTGGAGGCCCAGCTGCCCGTCATCAGAGTG
250 260 270 280 290 300 310 320
ACAGGTCTTATGACAGCCTGATTGGTGACTCGGGCTGGGTGTGGATTCTCACCCCAGGCCTCTGCCTGCTTTCTCAGACC
330 340 350 360 370 380 390 400
CTCATCTGTCACCCCCACGCTGAACCCAGCTGCCACCCCCAGAAGCCCATCAGACTGCCCCCAGCACACGGAATGGATTT
410 420 430 440 450 460 470 480
CTGAGAAAGAAGCCGAAACAGAAGGCCCGTGGGAGTCAGCATGCCGCGTGGCTGGGCCGCCCCCTTGCTCCTGCTGCTGC
M P R G W A A P L L L L L L
490 500 510 520 530 540 550 560
TCCAGGGAGGCTGGGGCTGCCCCGACCTCGTCTGCTACACCGATTACCTCCAGACGGTCATCTGCATCCTGGAAATGTGG
Q G G W G C P D L V C Y T D Y L Q T V I C I L E M W
570 580 590 600 610 620 630 640
AACCTCCACCCCAGCACGCTCACCCTTACCTGGCAAGACCAGTATGAAGAGCTGAAGGACGAGGCCACCTCCTGCAGCCT
N L H P S T L T L T W Q D Q Y E E L K D E A T S C S L
650 660 670 680 690 700 710 720
CCACAGGTCGGCCCACAATGCCACGCATGCCACCTACACCTGCCACATGGATGTATTCCACTTCATGGCCGACGACATTT
H R S A H N A T H A T Y T C H M D V F H F M A D D I F
730 740 750 760 770 780 790 800
TCAGTGTCAACATCACAGACCAGTCTGGCAACTACTCCCAGGAGTGTGGCAGCTTTCTCCTGGCTGAGAGCATCAAGCCG
S V N I T D Q S G N Y S Q E C G S F L L A E S I K P
810 820 830 840 850 860 870 880
GCTCCCCCTTTCAACGTGACTGTGACCTTCTCAGGACAGTATAATATCTCCTGGCGCTCAGATTACGAAGACCCTGCCTT
A P P F N V T V T F S G Q Y N I S W R S D Y E D P A F
890 900 910 920 930 940 950 960
CTACATGCTGAAGGGCAAGCTTCAGTATGAGCTGCAGTACAGGAACCGGGGAGACCCCTGGGCTGTGAGTCCGAGGAGAA
Y M L K G K L Q Y E L Q Y R N R G D P W A V S P R R K
970 980 990 1000 1010 1020 1030 1040
AGCTGATCTCAGTGGACTCAAGAAGTGTCTCCCTCCTCCCCCTGGAGTTCCGCAAAGACTCGAGCTATGAGCTGCAGGTG
L I S V D S R S V S L L P L E F R K D S S Y E L Q V
1050 1060 1070 1080 1090 1100 1110 1120
CGGGCAGGGCCCATGCCTGGCTCCTCCTACCAGGGGACCTGGAGTGAATGGAGTGACCCGGTCATCTTTCAGACCCAGTC
R A G P M P G S S Y Q G T W S E W S D P V I F Q T Q S
1130 1140 1150 1160 1170 1180 1190 1200
AGAGGAGTTAAAGGAAGGCTGGAACCCTCACCTGCTGCTTCTCCTCCTGCTTGTCATAGTCTTCATTCCTGCCTTCTGGA

[Fig. 5-2]



Note) The arrows show the positions of primers used for RT-PCR. They are, SN1 (798-827), SN2 (894-923), AS2 (1055-1026), and AS1 (1127-1098) from the 5′ side, in their order. For two bases at the 5′ end of AS1, AC, which is derived from the genomic sequence, was used in place of CT.

[Fig. 6-1]

```
        10         20         30         40         50         60         70         80
GGCAGCCAGCGGCCTCAGACAGACCCACTGGCGTCTCTCTGCTGAGTGACCGTAAGCTCGGCGTCTGGCCCTCTGCCTGC

        90        100        110        120        130        140        150        160
CTCTCCCTGAGTGTGGCTGACAGCCACGCAGCTGTGTCTGTCTGTCTGCGGCCCGTGCATCCCTGCTGCGGCCGCCTGGT

       170        180        190        200        210        220        230        240
ACCTTCCTTGCCGTCTCTTTCCTCTGTCTGCTGCTCTGTGGGACACCTGCCTGGAGGCCCAGCTGCCCGTCATCAGAGTG

       250        260        270        280        290        300        310        320
ACAGGTCTTATGACAGCCTGATTGGTGACTCGGGCTGGGTGTGGATTCTCACCCCAGGCCTCTGCCTGCTTTCTCAGACC

       330        340        350        360        370        380        390        400
CTCATCTGTCACCCCCACGCTGAACCCAGCTGCCACCCCCAGAAGCCCATCAGACTGCCCCCAGCACACGGAATGGATTT

       410        420        430        440        450        460        470        480
CTGAGAAAGAAGCCGAAACAGAAGGCCCGTGGGAGTCAGCATGCCGCGTGGCTGGGCCGCCCCCTTGCTCCTGCTGCTGC
                                              M  P  R  G  W  A  A  P  L  L  L  L  L  L

       490        500        510        520        530        540        550        560
TCCAGGGAGGCTGGGGCTGCCCCGACCTCGTCTGCTACACCGATTACCTCCAGACGGTCATCTGCATCCTGGAAATGTGG
  Q  G  G  W  G  C  P  D  L  V  C  Y  T  D  Y  L  Q  T  V  I  C  I  L  E  M  W

       570        580        590        600        610        620        630        640
AACCTCCACCCCAGCACGCTCACCCTTACCTGGCAAGACCAGTATGAAGAGCTGAAGGACGAGGCCACCTCCTGCAGCCT
N  L  H  P  S  T  L  T  L  T  W  Q  D  Q  Y  E  E  L  K  D  E  A  T  S  C  S  L

       650        660        670        680        690        700        710        720
CCACAGGTCGGCCCACAATGCCACGCATGCCACCTACACCTGCCACATGGATGTATTCCACTTCATGGCCGACGACATTT
 H  R  S  A  H  N  A  T  H  A  T  Y  T  C  H  M  D  V  F  H  F  M  A  D  D  I  F
                M  P  R  M  P  P  T  P  A  T  W  M  Y  S  T  S  W  P  T  T  F
       730        740        750        760        770        780        790        800
TCAGTGTCAACATCACAGACCAGTCTGGCAACTACTCCCAGGAGTGTGGCAGCTTTCTCCTGGCTGAGAGCAAGTCCGAG
  S  V  N  I  T  D  Q  S  G  N  Y  S  Q  E  C  G  S  F  L  L  A  E  S  K  S  E
S  V  S  T  S  Q  T  S  L  A  T  T  P  R  S  V  A  A  F  S  W  L  R  A  S  P  R
       810        820        830        840        850        860        870        880
GAGAAAGCTGATCTCAGTGGACTCAAGAAGTGTCTCCCTCCTCCCCCTGGAGTTCCGCAAAGACTCGAGCTATGAGCTGC
E  K  A  D  L  S  G  L  K  K  C  L  P  P  P  P  G  V  P  Q  R  L  E  L  *
 R  K  L  I  S  V  D  S  R  S  V  S  L  L  P  L  E  F  R  K  D  S  S  Y  E  L  Q
       890        900        910        920        930        940        950        960
AGGTGCGGGCAGGGCCCATGCCTGGCTCCTCCTACCAGGGGACCTGGAGTGAATGGAGTGACCCGGTCATCTTTCAGACC

  V  R  A  G  P  M  P  G  S  S  Y  Q  G  T  W  S  E  W  S  D  P  V  I  F  Q  T
```

[Fig. 6-2]

```
      970       980       990      1000      1010      1020      1030      1040
CAGTCAGAGGAGTTAAAGGAAGGCTGGAACCCTCACCTGCTGCTTCTCCTCCTGCTTGTCATAGTCTTCATTCCTGCCTT

Q  S  E  E  L  K  E  G  W  N  P  H  L  L  L  L  L  L  L  V  I  V  F  I  P  A  F
     1050      1060      1070      1080      1090      1100      1110      1120
CTGGAGCCTGAAGACCCATCCATTGTGGAGGCTATGGAAGAAGATATGGGCCGTCCCCAGCCCTGAGCGGTTCTTCATGC

 W  S  L  K  T  H  P  L  W  R  L  W  K  K  I  W  A  V  P  S  P  E  R  F  F  M  P
     1130      1140      1150      1160      1170      1180      1190      1200
CCCTGTACAAGGGCTGCAGCGGAGACTTCAAGAAATGGGTGGGTGCACCCTTCACTGGCTCCAGCCTGGAGCTGGGACCC

  L  Y  K  G  C  S  G  D  F  K  K  W  V  G  A  P  F  T  G  S  S  L  E  L  G  P
     1210      1220      1230      1240      1250      1260      1270      1280
TGGAGCCCAGAGGTGCCCTCCACCCTGGAGGTGTACAGCTGCCACCCACCCAGCAGCCCTGTGGAGTGTGACTTCACCAG

W  S  P  E  V  P  S  T  L  E  V  Y  S  C  H  P  P  S  S  P  V  E  C  D  F  T  S
     1290      1300      1310      1320      1330      1340      1350      1360
CCCCGGGGACGAAGGACCCCCCCGGAGCTACCTCCGCCAGTGGGTGGTCATTCCTCCGCCACTTTCGAGCCCTGGACCCC

 P  G  D  E  G  P  P  R  S  Y  L  R  Q  W  V  V  I  P  P  P  L  S  S  P  G  P  Q

     1370      1380      1390      1400      1410      1420      1430      1440
AGGCCAGCTAATGAGGCTGACTGGATGTCCAGAGCTGGCCAGGCCACTGGGCCCTGAGCCAGAGACAAGGTCACCTGGGC

  A  S  *  *

     1450      1460      1470      1480      1490      1500      1510      1520
TGTGATGTGAAGACACCTGCAGCCTTTGGTCTCCTGGATGGGCCTTTGAGCCTGATGTTTACAGTGTCTGTGTGTGTGTG


     1530      1540      1550      1560      1570      1580      1590      1600
CATATGTGTGTGTGTGCATATGCATGTGTGTGTGTGTGTGTGTCTTAGGTGCGCAGTGGCATGTCCACGTGTGTGTGATT


     1610      1620      1630      1640      1650      1660      1670      1680
GCACGTGCCTGTGGGCCTGGGATAATGCCCATGGTACTCCATGCATTCACCTGCCCTGTGCATGTCTGGACTCACGGAGC


     1690      1700      1710      1720      1730      1740      1750      1760
TCACCCATGTGCACAAGTGTGCACAGTAAACGTGTTTGTGGTCAACAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA


     1770      1780
AAAAAAAAAAAAAAAAAA
```

Note) Two possible open reading frames (ORF) are shown.

[Fig. 7-1]

```
        10         20         30         40         50         60         70         80
GGCAGCCAGCGGCCTCAGACAGACCCACTGGCGTCTCTCTGCTGAGTGACCGTAAGCTCGGCGTCTGGCCCTCTGCCTGC

        90        100        110        120        130        140        150        160
CTCTCCCTGAGTGTGGCTGACAGCCACGCAGCTGTGTCTGTCTGTCTGCGGCCCGTGCATCCCTGCTGCGGCCGCCTGGT

       170        180        190        200        210        220        230        240
ACCTTCCTTGCCGTCTCTTTCCTCTGTCTGCTGCTCTGTGGGACACCTGCCTGGAGGCCCAGCTGCCCGTCATCAGAGTG

       250        260        270        280        290        300        310        320
ACAGGTCTTATGACAGCCTGATTGGTGACTCGGGCTGGGTGTGGATTCTCACCCCAGGCCTCTGCCTGCTTTCTCAGACC

       330        340        350        360        370        380        390        400
CTCATCTGTCACCCCCACGCTGAACCCAGCTGCCACCCCCAGAAGCCCATCAGACTGCCCCCAGCACACGGAATGGATTT

       410        420        430        440        450        460        470        480
CTGAGAAAGAAGCCGAAACAGAAGGCCCGTGGGAGTCAGCATGCCGCGTGGCTGGGCCGCCCCCTTGCTCCTGCTGCTGC
                                              M  P  R  G  W  A  A  P  L  L  L  L  L  L
       490        500        510        520        530        540        550        560
TCCAGGGAGGCTGGGGCTGCCCCGACCTCGTCTGCTACACCGATTACCTCCAGACGGTCATCTGCATCCTGGAAATGTGG
  Q  G  G  W  G  C  P  D  L  V  C  Y  T  D  Y  L  Q  T  V  I  C  I  L  E  M  W
       570        580        590        600        610        620        630        640
AACCTCCACCCCAGCACGCTCACCCTTACCTGGCAAGACCAGTATGAAGAGCTGAAGGACGAGGCCACCTCCTGCAGCCT
N  L  H  P  S  T  L  T  L  T  W  Q  D  Q  Y  E  E  L  K  D  E  A  T  S  C  S  L
       650        660        670        680        690        700        710        720
CCACAGGTCGGCCCACAATGCCACGCATGCCACCTACACCTGCCACATGGATGTATTCCACTTCATGGCCGACGACATTT
 H  R  S  A  H  N  A  T  H  A  T  Y  T  C  H  M  D  V  F  H  F  M  A  D  D  I  F
       730        740        750        760        770        780        790        800
TCAGTGTCAACATCACAGACCAGTCTGGCAACTACTCCCAGGAGTGTGGCAGCTTTCTCCTGGCTGAGAGCATCAAGCCG
  S  V  N  I  T  D  Q  S  G  N  Y  S  Q  E  C  G  S  F  L  L  A  E  S  I  K  P
       810        820        830        840        850        860        870        880
GCTCCCCCTTTCAACGTGACTGTGACCTTCTCAGGACAGTATAATATCTCCTGGCGCTCAGATTACGAAGACCCTGCCTT
A  P  P  F  N  V  T  V  T  F  S  G  Q  Y  N  I  S  W  R  S  D  Y  E  D  P  A  F
       890        900        910        920        930        940        950        960
CTACATGCTGAAGGGCAAGCTTCAGTATGAGCTGCAGTACAGGAACCGGGGAGACCCCTGGGCTGTGAGTCCGAGGAGAA
 Y  M  L  K  G  K  L  Q  Y  E  L  Q  Y  R  N  R  G  D  P  W  A  V  S  P  R  R  K
       970        980        990       1000       1010       1020       1030       1040
AGCTGATCTCAGTGGACTCAAGAAGTGTCTCCCTCCTCCCCCTGGAGTTCCGCAAAGACTCGAGCTATGAGCTGCAGGTG
  L  I  S  V  D  S  R  S  V  S  L  L  P  L  E  F  R  K  D  S  S  Y  E  L  Q  V
      1050       1060       1070       1080       1090       1100       1110       1120
CGGGCAGGGCCCATGCCTGGCTCCTCCTACCAGGGGACCTGGAGTGAATGGAGTGACCCGGTCATCTTTCAGACCCAGTC
R  A  G  P  M  P  G  S  S  Y  Q  G  T  W  S  E  W  S  D  P  V  I  F  Q  T  Q  S
      1130       1140       1150       1160       1170       1180       1190       1200
AGAGGAGTTAAAGGAAGGCTGGAACCCTCACCTGCTGCTTCTCCTCCTGCTTGTCATAGTCTTCATTCCTGCCTTCTGGA
```

[Fig. 7-2]

```
 E  E  L  K  E  G  W  N  P  H  L  L  L  L  L  L  V  I  V  F  I  P  A  F  W  S
      1210       1220       1230       1240       1250       1260       1270       1280
GCCTGAAGACCCATCCATTGTGGAGGCTATGGAAGAAGATATGGGCCGTCCCCAGCCCTGAGCGGTTCTTCATGCCCCTG
  L  K  T  H  P  L  W  R  L  W  K  K  I  W  A  V  P  S  P  E  R  F  F  M  P  L
      1290       1300       1310       1320       1330       1340       1350       1360
TACAAGGGCTGCAGCGGAGACTTCAAGAAATGGGTGGGTGCACCCTTCACTGGCTCCAGCCTGGAGCTGGGACCCTGGAG
Y  K  G  C  S  G  D  F  K  K  W  V  G  A  P  F  T  G  S  S  L  E  L  G  P  W  S
      1370       1380       1390       1400       1410       1420       1430       1440
CCCAGAGGTGCCCTCCACCCTGGAGGTGTACAGCTGCCACCCACCACGGAGCCCGGCCAAGAGGCTGCAGCTCACGGAGC
 P  E  V  P  S  T  L  E  V  Y  S  C  H  P  P  R  S  P  A  K  R  L  Q  L  T  E  L
      1450       1460       1470       1480       1490       1500       1510       1520
TACAAGAACCAGCAGAGCTGGTGGAGTCTGACGGTGTGCCCAAGCCCAGCTTCTGGCCGACAGCCCAGAACTCGGGGGGC
  Q  E  P  A  E  L  V  E  S  D  G  V  P  K  P  S  F  W  P  T  A  Q  N  S  G  G

      1530       1540       1550       1560       1570       1580       1590       1600
TCAGCTTACAGTGAGGAGAGGGATCGGCCATACGGCCTGGTGTCCATTGACACAGTGACTGTGCTAGATGCAGAGGGGCC
 S  A  Y  S  E  E  R  D  R  P  Y  G  L  V  S  I  D  T  V  T  V  L  D  A  E  G  P

      1610       1620       1630       1640       1650       1660       1670       1680
ATGCACCTGGCCCTGCAGCTGTGAGGATGACGGCTACCCAGCCCTGGACCTGGATGCTGGCCTGGAGCCCAGCCCAGGCC
 C  T  W  P  C  S  C  E  D  D  G  Y  P  A  L  D  L  D  A  G  L  E  P  S  P  G  L

      1690       1700       1710       1720       1730       1740       1750       1760
TAGAGGACCCACTCTTGGATGCAGGGACCACAGTCCTGTCCTGTGGCTGTGTCTCAGCTGGCAGCCCTGGGCTAGGAGGG
  E  D  P  L  L  D  A  G  T  T  V  L  S  C  G  C  V  S  A  G  S  P  G  L  G  G

      1770       1780       1790       1800       1810       1820       1830       1840
CCCCTGGGAAGCCTCCTGGACAGACTAAAGCCACCCCTTGCAGATGGGGAGGACTGGGCTGGGGGACTGCCCTGGGGTGG
 P  L  G  S  L  L  D  R  L  K  P  P  L  A  D  G  E  D  W  A  G  G  L  P  W  G  G

      1850       1860       1870       1880       1890       1900       1910       1920
CCGGTCACCTGGAGGGGTCTCAGAGAGTGAGGCGGGCTCACCCCTGGCCGGCCTGGATATGGACACGTTTGACAGTGGCT
 R  S  P  G  G  V  S  E  S  E  A  G  S  P  L  A  G  L  D  M  D  T  F  D  S  G  F

      1930       1940       1950       1960       1970       1980       1990       2000
TTGTGGGCTCTGACTGCAGCAGCCCTGTGGAGTGTGACTTCACCAGCCCCGGGGACGAAGGACCCCCCCGGAGCTACCTC
  V  G  S  D  C  S  S  P  V  E  C  D  F  T  S  P  G  D  E  G  P  P  R  S  Y  L

      2010       2020       2030       2040       2050       2060       2070       2080
CGCCAGTGGGTGGTCATTCCTCCGCCACTTTCGAGCCCTGGACCCCAGGCCAGCTAATGAGGCTGACTGGATGTCCAGAG
R  Q  W  V  V  I  P  P  P  L  S  S  P  G  P  Q  A  S  *  *

      2090       2100       2110       2120       2130       2140       2150       2160
CTGGCCAGGCCACTGGGCCCTGAGCCAGAGACAAGGTCACCTGGGCTGTGATGTGAAGACACCTGCAGCCTTTGGTCTCC

      2170       2180       2190       2200       2210       2220       2230       2240
```

[Fig. 7-3]

TGGATGGGCCTTTGAGCCTGATGTTTACAGTGTCTGTGTGTGTGTGCATATGTGTGTGTGTGCATATGCATGTGTGTGTG

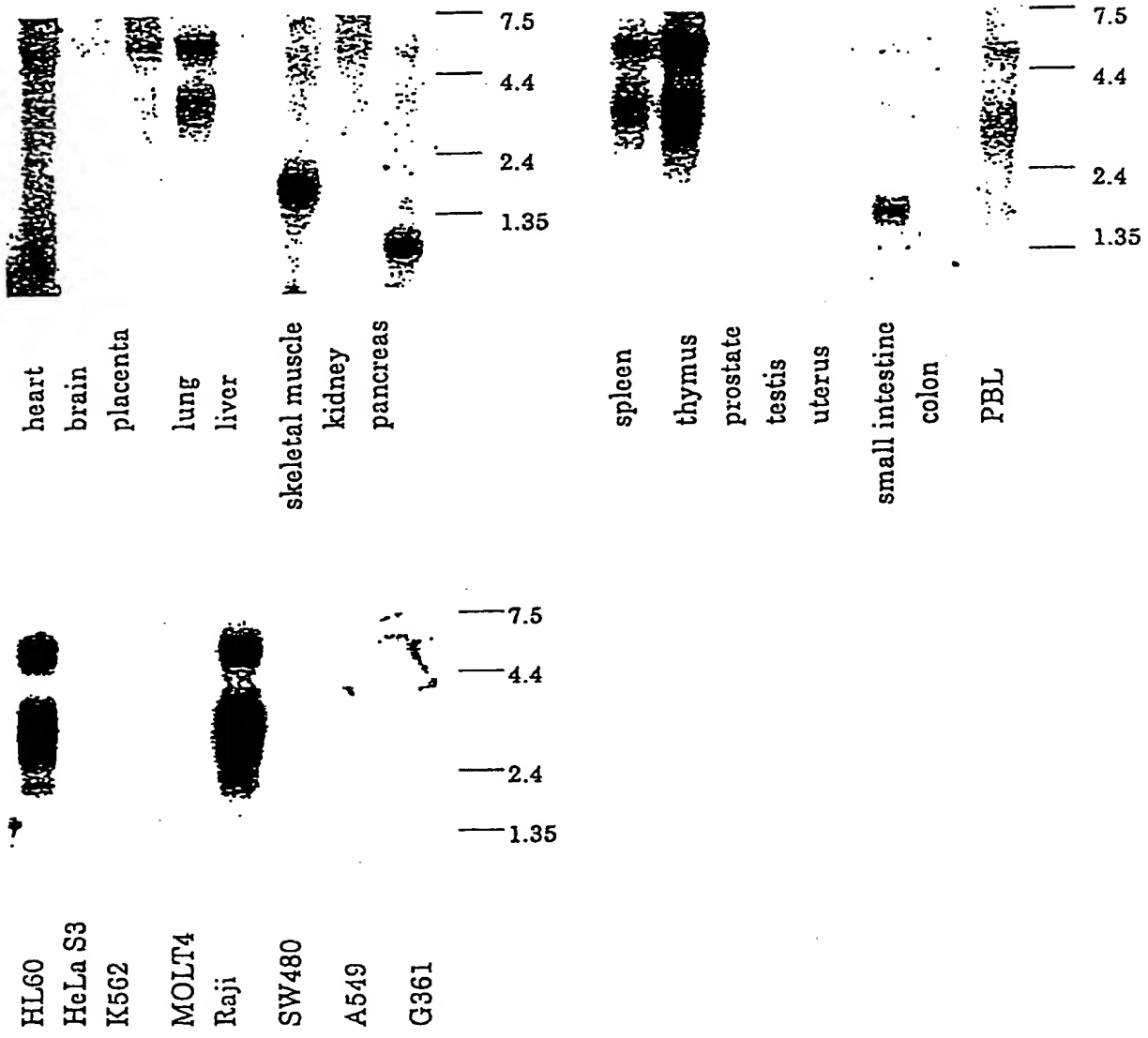
2250 2260 2270 2280 2290 2300 2310 2320
TGTGTGTGTCTTAGGTGCGCAGTGGCATGTCCACGTGTGTGTGATTGCACGTGCCTGTGGGCCTGGGATAATGCCCATGG

2330 2340 2350 2360 2370 2380 2390 2400
TACTCCATGCATTCACCTGCCCTGTGCATGTCTGGACTCACGGAGCTCACCCATGTGCACAAGTGTGCACAGTAAACGTG

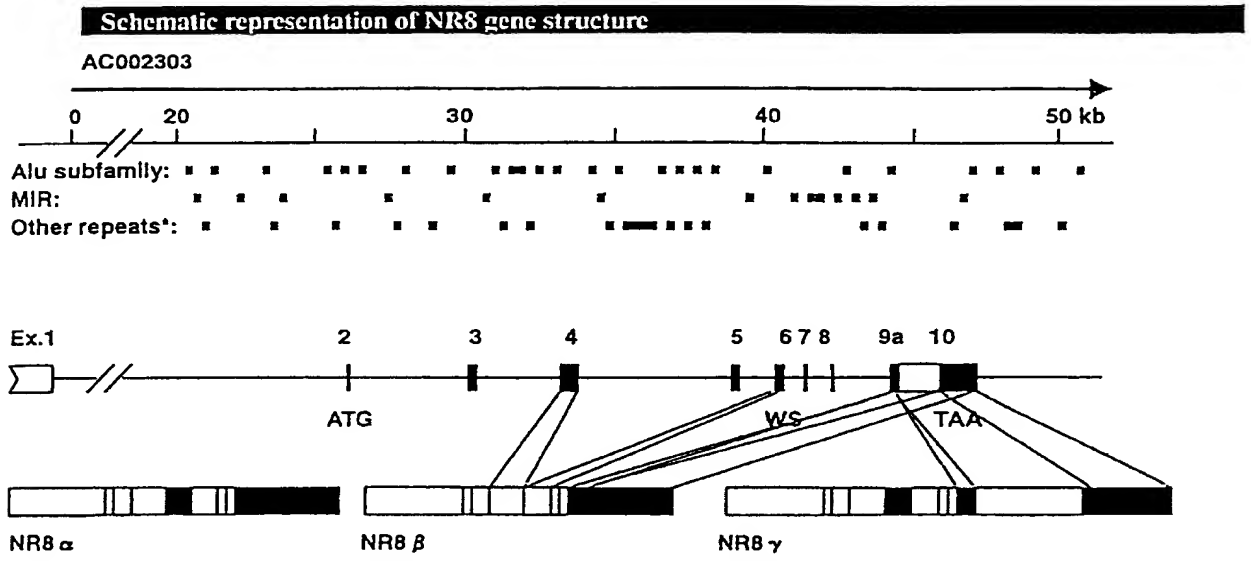
2410 2420 2430 2440 2450 2460 2470 2480
TTTGTGGTCAACAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Note) The 177 amino acids inserted by alternative splicing are underlined.

[Fig. 8]


7.5
4.4
2.4
1.35
heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas
spleen
thymus
prostate
testis
uterus
small intestine
colon
PBL
HL60
HeLa S3
K562
MOLT4
Raji
SW480
A549
G361

[Fig. 9]



* Other repeats include (CA)n, (CAGA)n, (TGGA)n, (CATA)n, (TA)n, (GA)n, (GGAA)n, (CATG)n, (GAAA)n, MSTA, AT-rich, MLT1A1, LINE2, FLAM_C, MER63A, MSTB.

[Table 1-1]

Table 1. Result of 2 steps Blast search

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TGGAGTAATTGGAGC	Asn	AL009181	30692 tggagtaattggagc 30678	1p34.1-1p35	mIL11R(opposite), GOBP
TGGAGCTGATGGAGC	***	Z97987	140006 tggagctgatggagc 139992	1p36.2-36.3	line1, Leu Zip p40,
TGGAGCAGCTGGAGC	Ser	AF023268	39931 tggagcagctggagc 39917	1q21	metaxin
TGGAGCTGCTGGAGC	Cys	AL009051	78023 tggagctgctggagc 78037	1q23-24	HP-10, semaphorin F,G
TGGAGCACGTGGAGT	Thr	Z97200	112905 tggagcacgtggagt 112891	1q24	AFP enhancer BP, RAR
TGGAGTGCCTGGAGC	Ala	U95626	101031 tggagtgcctggagc 101017	3	CFTC, TcR
TGGAGTAGATGGAGT	Arg	Z84495	2547 tggagtagatggagt 2533	3p21.3	trithorax
TGGAGCTGATGGAGT	***	Z74023	5255 tggagctgatggagt 5241	3p21.3	E2ABP, fibronectin, nidgen
TGGAGTTTCTGGAGT	Phe	Z68275	7291 tggagtttctggagt 7277	4p16.3	mena, NMDAR
TGGAGTGCCTGGAGT	Ala	Z54072	21277 tggagtgcctggagt 21291	4p16.3	crk, AchR, HER3
TGGAGCTGCTGGAGC	Cys	Z69837	30266 tggagctgctggagc 30252	4p16.3	KIT, FLT3, PDGFRa
TGGAGTTACTGGAGT	Tyr	AC003951	27290 tggagttactggagt 27304	5	collagen
TGGAGCCTGTGGAGT	Leu	AC004502	48334 tggagcctgtggagt 48320	5	ADAMTS-1, properdin, etc
TGGAGTTGATGGAGC	***	L81613	2418 tggagttgatggagc 2404	5	APC, bat2, p53
TGGAGTGTATGGAGT	Val	AC002122	43679 tggagtgtatggagt 43665	5p15.2	Met tRNA syntase
TGGAGTCCATGGAGT	Pro	AC002380	34646 tggagtccatggagt 34632	5p15.2	N-WASP, enigma
TGGAGCAACTGGAGC	Asn	AC002479	80443 tggagcaactggagc 80457	5p15.2	NEU, glycoprotein C
TGGAGCTGCTGGAGT	Cys	AC004592	125445 tggagctgctggagt 125431	5q31	CD22-B
TGGAGTAGCTGGAGT	Ser	AC002393	3721 tggagtagctggagt 3735	6	glycoprotein
TGGAGTTGCTGGAGT	Cys	AC002326	114578 tggagttgctggagt 114564	6	G3P REGULON
TGGAGTGCATGGAGT	Ala	Z84490	20244 tggagtgcatggagt 20230	6	Alu, adrenergic receptor

[Table 1-2]

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TTGGAGTTTCTGGAGC	Phe	AC002112	68699 tggagtttctggagc 68685	6	IgHv, MYD116
TGGAGCGGCTGGAGC	Gly	U89336	35829 tggagcggctggagc 35815	6p21	myosin HC, cep250,
TGGAGCGTCTGGAGC	Val	U53588	3558 tggagcgtctggagc 3572	6p21.3	ring finger, BRCA1
TGGAGTGCATGGAGT	Ala	Z98744	38358 tggagtgcatggagt 38344	6p21.3-22.3	Alu, AD7c-NTP
TGGAGTTGCTGGAGT	Cys	AL009031	104325 tggagttgctggagt 104311	6p22.3-24.1	ACC synthase
TGGAGTGTCTGGAGT	Val	AL008729	21325 tggagtgtctggagt 21339	6p24	E1A, DUB-2
TGGAGTTGTTGGAGT	Cys	Z98755	69825 tggagttgttggagt 69811	6q16.1-21	dynein
TGGAGCTTCTGGAGC	Phe	Z98172	35554 tggagcttctggagc 35540	6q21	HGXPRT
TGGAGCAGGTGGAGC	Arg	Z97989	79116 tggagcaggtggagc 79102	6q21-22	syn fyn, slk, yes, src
TGGAGCTAATGGAGT	***	Z95326	16562 tggagctaatggagt 16576	6q22.1-6q22.33	tyrosinase
TGGAGCTCTTGGAGC	Ser	Z98049	25800 tggagctcttggagc 25786	6q26-q27	collagen, AT3, C1Qb
TGGAGCTCCTGGAGT	Ser	AC003090	22068 tggagctcctggagt 22082	7p15	ICE
TGGAGTATATGGAGC	Ile	AC004744	22740 tggagtatatggagc 22754	7p15-p21	TSH-R, RNABP
TGGAGTAGCTGGAGC	Ser	AC004485	86356 tggagtagctggagc 86370	7p15-p21	Hox 2.4, mIL11Ra(stop*)
TGGAGTCTTTGGAGT	Leu	AC004141	3130 tggagtctttggagt 3144	7p21-p22	polyprotein
TGGAGCAGATGGAGC	Arg	AC004548	62876 tggagcagatggagc 62862	7q11.23-q21.1	NCAM
TGGAGCAACTGGAGT	Asn	AC002456	69500 tggagcaactggagt 69514	7q21	glycoprotein A
TGGAGTAACTGGAGT	Asn	AC000064	9170 tggagtaactggagt 9184	7q21-22	GA3PD
TGGAGTTATTGGAGT	Tyr	AC003085	87341 tggagttattggagt 87355	7q21-22	Nmyc, FGFR
TGGAGTTGTTGGAGT	Cys	AC000119	65235 tggagttgttggagt 65221	7q21-7q22	FVIII, TopoIII
TGGAGTTGTTGGAGT	Cys	AC002458	44435 tggagttgttggagt 44421	7q21-q22	telomerase, NFAT
TGGAGTACATGGAGC	Thr	AC000059	9977 tggagtacatggagc 9963	7q21-7q22	Alu, Notch4

[Table 2]

Table 2 NR8 CDS on Chromosome 16p12(AC002303)

Exon	# in AC002303	# in NR8	Features
1	<1	: 1-424	in frame stop codon
2	26334-26398	: 425-489	initiation codon, signal peptide
3	30625-30727	: 490-592	conserved Cys residues
4	33766-33965	: 593-792	conserved Cys residues, N-glycosylation sites
5	39240-39394	: 793-947	Pro-rich motif (PAPPF), N-glycosylation sites
6	40820-40997	: 948-1125	gtWSEWSdp motif
7	41455-41554	: 1126-1225	transmembrane domain
8	42285-42366	: 1226-1307	Box1 (IWAVPSP)
9a	44812-44909	: 1308-1405*	join to exon10, Box2? (PSTLEVYSCH), non-conserved boundary
9b	44812-45922<	: 1308-2465**	double stop codons, Box2? (PSTLEVYSCH, PAELVESDG), poly A
10	45441-45922<	: 1406-1934*	double stop codons, poly A

NR8 alpha* : Exons 1+2÷3+4+5+6÷7+8+9a+10

NR8 beta : Exons 1+2+3+4+6+7+8÷9a+10(two alternative reading frames for soluble and TM (-signal) forms)

NR8 gamma** : Exons 1+2+3+4+5+6+7+8+9b (hypothetical)

[Table 3]

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TGGAGTCAGTGGAGC	Gln	AC000...	...tggagtcagtggagc...	7q21-7q22	reverse transcriptase
TTGGAGTATTTGGAGT	Ile	AC002384	52216 tggagtatttggagt 52202	7q22	pol, GHR(another frame)
TGGAGCAGCTGGAGT	Ser	AC004522	55291 tggagcagctggagt 55277	7q22-q31.1	hemoglobin beta
TGGAGTGTTTGGAGT	Val	AC002466	43273 tggagtgtttggagt 43287	7q31	ryanodine receptor、mTPO
TGGAGTGGCTGGAGC	Gly	AC002543	112948 tggagtggctggagc 112962	7q31.2	EGF, P-selectin
TGGAGCTGATGGAGC	***	AC000061	79564 tggagctgatggagc 79550	7q31.2	laminin B1, tubulin
TGGAGTTTTTGGAGT	Phe	AC000125	13750 tggagttttggagt 13736	7q31.3	p150
TGGAGTTGTTGGAGT	Cys	AC002498	20166 tggagttgttggagt 20152	7q31.3	IL3Rb(opposite)
TGGAGCGGGTGGAGC	Gly	U66059	158491 tggagcgggtggagc 158477	7q35(TcRb)	properdin
TGGAGCATTTGGAGC	Ile	AC003109	4761 tggagcatttggagc 4775	7q36	CD2, HOX-2.6
TGGAGTTATTGGAGT	Tyr	AF027390	174448 tggagttattggagt 174434	7q tel	IkB, V2R
TGGAGCATATGGAGT	Ile	AC002052	28882 tggagcatatggagt 28896	9p22	myosin VIIA, OSMRb
TGGAGCAACTGGAGT	Asn	AC001643	27345 tggagcaactggagt 27331	9q34	hox1.4, gastrinR
...	Asn	AC002319	37518 tggagcaactggagt 37532	9q34	hox1.4, gastrinR
TGGAGCGGATGGAGC	Gly	AC000396	16394 tggagcggatggagc 16380	9q34	vWf, laminin a3
TGGAGTGAGTGGAGT	Glu	U73649	16850 tggagtgagtggagt 16836	11	zinc finger
TGGAGTGGGTGGAGT	Gly	U73649	16859 tggagtgggtggagt 16845	11	zinc finger
TGGAGTGCCTGGAGT	Ala	U73629	31027 tggagtgcctggagt 31041	11	Alu, gp2b, BCGF-12
TGGAGCCCCTGGAGT	Pro	U73629	36731 tggagccccctggagt 36745	11	EZRIN, hemagglutinin
TGGAGTCCCTGGAGC	Pro	U73643	14550 tggagtccctggagc 14564	11	reverse transcriptase
TGGAGCAACTGGAGC	Asn	AF015416	65621 tggagcaactggagc 65635	11p15.5	Nasopressin R, OSNR
TGGAGTGCATGGAGT	Ala	AC002350	23543 tggagtgcatggagt 23529	12q24	Alu, IFNaR

[Table 4]

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TGGAGTGCATGGAGT	Ala	AC004217	88822 tggagtgcatggagt 88808	12q24.1	Alu, HPK
TTGGAGTTACTGGAGC	Tyr	AC002978	65893 tggagttactggagc 65907	12q24	clathrin LC, EPOR(nonWS)
TGGAGTTGTTGGAGT	Cys	AC000403	91715 tggagttgttggagt 91729	13	VHL, inhibin B
TGGAGCGGTTGGAGC	Gly	X97051	73621 tggagcggttggagc 73607	14q32.33 (IgD)	polycystic kidney
TGGAGTAGGTGGAGC	Arg	AC003024	15596 tggagtaggtggagc 15582	15q26	pksF
TGGAGTTTCTGGAGC	Phe	AC002492	93356 tggagtttctggagc 93370	16	pol, PRLR
TGGAGTTCATGGAGT	Ser	U91318	102406 tggagttcatggagt 102392	16	ICAM1, MIBP1
TGGAGTGTATGGAGT	Val	AC002289	10631 tggagtgtatggagt 10645	16	Alu
TGGAGCCACTGGAGT	His	U91318	152252 tggagccactggagt 152238	16	laminin alpha5
TGGAGTTAATGGAGT	***	AC002519	81768 tggagttaatggagt 81754	16	Rho, Notch
TGGAGCTGCTGGAGT	Cys	U91326	84127 tggagctgctggagt 84113	16p11.2	NIPI-like, IL2Rr(nonWS)
TGGAGTGAATGGAGT	Glu	AC002303	40952 tggagtgaatggagt 40966	16p12	TPOR, OBR, and many
TGGAGCACTTGGAGC	Thr	AC002551	82245 tggagcacttggagc 82259	16p12.1	envelope, androgen R
TGGAGTCCCTGGAGC	Pro	AC002299	162 tggagtcccctggagc 148	16p12-p13.1	CYCLIN H, FN
TGGAGCTATTGGAGC	Tyr	AC002299	84540 tggagctattggagc 84526	16p12.5-p13.1	Alu-RNA-editase
TGGAGTCACTGGAGT	His	U95737	16130 tggagtcactggagt 16144	16p13.1	TcRa, HLAa
			16374 tggagtcactggagt 16388		Notch, Pro-rich
			16599 tggagtcactggagt 16613		phosphatase, ORFB
TGGAGTCCTTGGAGC	Pro	U91318	112272 tggagtccttggagc 112286	16p13.13	CD80, collagen, MAP1a
TGGAGCACTTGGAGC	Thr	AC004509	26031 tggagcacttggagc 26045	16p13.3	TcRb
TGGAGCCGTTGGAGC	Arg	AC004496	28217 tggagccgttggagc 28231	16p13.3	mucin, ET1, IL12R(nonWS)

[Table 5]

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TGGAGCCGCTGGAGC	Arg	AC004232	34550 tggagccgctggagc 34564	16p13.3	IgLk, AGPR
TTGGAGTACTTGGAGC	Thr	AJ003147	151180 tggagtacttggagc 151166	16p13.3	RanBP2
TGGAGCGTGTGGAGC	Val	X71874	11520 tggagcgtgtggagc 11534	16q22.1	collagen a5IV
TGGAGCAAATGGAGT	Lys	AC003663	114346 tggagcaaatggagt 114360	17	beta-D-glucosidase
TGGAGTCTCTGGAGC	Leu	AC003957	52898 tggagtctctggagc 52884	17	TIE-1, SEX, Rho,
TGGAGCAGATGGAGC	Arg	AC003971	76277 tggagcagatggagc 76263	18	LIMK-1, TcR
TGGAGTGCATGGAGT	Ala	AD000812	30891 tggagtgcatggagt 30905	19	Alu
[illegible]	Ala	AC002126	85832tggagtgcatggagt85846	19	Alu, AD7c-NTP
TGGAGCTGCTGGAGT	Cys	AC004660	10008 tggagctgctggagt 10022	19	Reps1
TGGAGCCCCTGGAGT	Pro	AC004490	14389 tggagccccctggagt 14403	19	mucin, ataxin-2, N-WASP
TGGAGTGAGTGGAGC	Glu	AC003112	18315 tggagtgagtggagc 18301	19p12(NR6)	TPOR, PRLR, OBR, etc
TGGAGCAGATGGAGC	Arg	AC004004	39010 tggagcagatggagc 38996	19p12	PRLR, IL12R, GM-CSFRb, IL11R(+stop codon) IL3Ra(weak, 22 nonWS)
		-----presumably a pseudogene-----			
			39177 tggagcagatggagc 39163		
TGGAGCACCTGGAGT	Thr	AD000685	21015 tggagcacctggagt 21001	19p13.1	GM-CSFRb(nonWS+stop)
TGGAGCTGATGGAGC	***	AC002115	37164 tggagctgatggagc 37178	19q13.1	Mpc2, Pro rich protein
TGGAGCCAGTGGAGC	Gln	M63796	7622 tggagccagtggagc 7636	19q13.3	NFCP, titin, Jagged 2
TGGAGTTACTGGAGT	Tyr	AC004505	31711 tggagttactggagt 31725	20	Gap junction
TGGAGTTGATGGAGC	***	Z93016	31093 tggagttgatggagc 31079	20q12-13.2	smaphorin F, GHS-R, JAK2
TGGAGTCAATGGAGT	Gln	L35677	579 tggagtcaatggagt 565	21(MX1)	GLI, IL9R, IL7R(nonWS)
TGGAGTGCCTGGAGT	Ala	AF039907	29892 tggagtgcctggagt 29906	21	IgV, Cyt.Oxidase
TGGAGTGTCTGGAGT	Val	AG000937	105 tggagtgtctggagt 91	21q	peroxidasin

[Table 6]

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TGGAGTAAATGGAGT	Lys	AP000034	28803 tggagtaaatggagt 28789	21q11.1	Na/Ca exchanger
TTGGAGTAGGTGGAGT	Arg	AP000039	24900 tggagtaggtggagt 24914	21q11.1	RNA polymerase
TGGAGTGAGTGGAGT	Glu	AP000035	21721 tggagtgagtggagt 21707	21q11.1	smaphorin F
TGGAGTGTCTGGAGT	Val	AG000038	26164 tggagtgtctggagt 26150	21q11.1	Glycoprotein
TGGAGTGTCTGGAGT	Val	AP000038	26164 tggagtgtctggagt 26150	21q11.1	PKC
TGGAGTGCCTGGAGT	Ala	AP000045	7204 tggagtgcctggagt 7218	21q11.1	IgV,
TGGAGCATTTGGAGC	Ile	AP000052	93726 tggagcatttggagc 93740	21q11.1	Ig H, TCF-3, CETP
TGGAGCCTCTGGAGC	Leu	AP000037	17581 tggagcctctggagc 17567	21q11.1	Alu, BCGF
TGGAGTGGGTGGAGT	Gly	AP000015	48480 tggagtgggtggagt 48494	21q22.2	TPO
TGGAGTGAGTGGAGT	Glu	Z97055	151632 tggagtgagtggagt 151618	22	semaphorin H, CD44
TGGAGCTGGTGGAGT	Trp	Z83856	8503 tggagctggtggagt 8489	22	ERF
TGGAGTGGGTGGAGT	Gly	Z95113	69325 tggagtgggtggagt 69311	22q11.2-qter	factor H
TGGAGTGCATGGAGT	Ala	Z93784	36348 tggagtgcatggagt 36362	22q11.2-qter	Alu, NF2
TGGAGCCTCTGGAGT	Leu	AC002308	130741 tggagcctctggagt 130727	22q11.2	collagen a1, Na channel
TGGAGTCCCTGGAGC	Pro	AC000086	40705 tggagtcccctggagc 40691	22q11.2	ADH, collagen
TGGAGCATCTGGAGC	Ile	L77569	21088 tggagcatctggagc 21074	22q11DiGeorge	clathrin heavy chain 2
TGGAGCATCTGGAGC	Ile	AC000072	24248 tggagcatctggagc 24262	22q11.2	clathrin heavy chain 2
TGGAGCAGCTGGAGC	Ser	AC000092	9817 tggagcagctggagc 9803	22q11.2	IgHv, PC binding
TGGAGCAACTGGAGC	Asn	Z95116	64481 tggagcaactggagc 64495	22q12.1	p150, IL4RQVSNVSF*)
TGGAGCTAGTGGAGC	***	AC003071	114780 tggagctagtggagc 114794	22q12.1-qter	FGFRb
TGGAGCCCTTGGAGC	Pro	Z80902	2675 tggagcccttggagc 2661	22q12-qter	collagen a1
TGGAGCTCTTGGAGT	Ser	Z79999	40825 tggagctcttggagt 40839	22q12-qter	collagen a1,

[Table 7]

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TGGAGCCATTGGAGT	His	Z81308	12575 tggagccattggagt 12561	22q12-qter	MYF-5, p53, INK4a
TGGAGCGAGTGGAGT	Glu	AL008637	85322 tggagcgagtggagt 85336	22q12.3-13.2	GM-CSFRb,IL3R, EPOR, etc
TTGGAGTGAGTGGAGT	Glu	U62317	77740 tggagtgagtggagt 77726	22q13	latrophilin-related
TGGAGTGCATGGAGT	Ala	Z98046	31082 tggagtgcatggagt 31068	22q13	Alu, G-CSFR, AD7c-NTP
TGGAGTTGTTGGAGT	Cys	AC002422	19151 tggagttgttggagt 19137	X	cGMP PDase
TGGAGTGTCTGGAGT	Val	Z73418	31830 tggagtgtctggagt 31816	X	WNT-8D, Mi-2
TGGAGTCTTTGGAGT	Leu	Z83843	114972 tggagtctttggagt 114958	X	reverse transcriptase
TGGAGTCTCTGGAGT	Leu	Z99706	7749 tggagtctctggagt 7735	X*	Selenoprotein
TGGAGCAACTGGAGT	Asn	AC002420	70704 tggagcaactggagt 70690	X	homeoprotein, OBR(stop)
TGGAGCATGTGGAGT	Met	Z77249	5702 tggagcatgtggagt 5688	X	TcRb, LIFR-α
TGGAGTTCCTGGAGC	Ser	Z83131	4904 tggagttcctggagc 4890	X	VPS41 homolog
TGGAGTGGCTGGAGC	Gly	AC002388	239975 tggagtggctggagc 239989	X	GAP, mLIFR(stop)
TGGAGTGGCTGGAGC	Gly	AC004478	73509 tggagtggctggagc 73495	X	RNase, mLIFR(stop)
TGGAGTCTATGGAGC	Leu	Z70050	9934 tggagtctatggagc 9948	X	complement C8, C7
TGGAGTCTATGGAGC	Leu	Z73986	40766 tggagtctatggagc 40780	X	complement C8, C7
TGGAGCTGTTGGAGC	Cys	L44140	112657 tggagctgttggagc 112671	X	rab GDI alpha, BDGF
TGGAGCTCATGGAGC	Ser	AC004383	144906 tggagctcatggagc 144892	X	RTase, transposon
TGGAGTAAATGGAGC	Lys	Z69732	31681 tggagtaaatggagc 31695	Xp11	OT-R, acrosin
TGGAGTTCGTGGAGC	Ser	Z92545	88703 tggagttcgtggagc 88717	Xp11	PMK1
TGGAGCTTCTGGAGC	Phe	AL008709	46089 tggagcttctggagc 46075	Xp11.23-Xp11.4	rMHC class 1a, HLA-C
TGGAGTTTCTGGAGT	Phe	U96409	116332 tggagtttctggagt 116346	Xp22	myosin H
TGGAGTTGCTGGAGT	Cys	AC003106	89544 tggagttgctggagt 89530	Xp22	IL9R

[Table 8]

Probe	Xaa	accession	location of hit	locus	blastx (expect=100)
TTGGAGTCACTGGAGT	His	AL021706	11982 tggagtcactggagt 11968	Xq21.1-21.33	dopamine receptor
TGGAGCTGGTGGAGT	Trp	AC000113	119188 tggagctggtggagt 119202	Xq23	DNA repair protein, MHC
TGGAGCAAGTGGAGT	Lys	AF007262	98212 tggagcaagtggagt 98226	Xq28	RNA polymerase
TGGAGCTGCTGGAGT	Cys	U82671	35792 tggagctgctggagt 35806	Xq28	XTCF-3c
TGGAGTCAGTGGAGC	Gln	AF011889	144465 tggagtcagtggagc 144451	Xq28	GHRHR, Werner Synd.
TGGAGCTAATGGAGC	***	AF030876	107409 tggagctaatggagc 107395	Xq28	gp41, clk3
[redacted]		[redacted]	[redacted]	[redacted]	[redacted]
[redacted]		[redacted]	[redacted]	[redacted]	[redacted]
TGGAGTTTCTGGAGT	Phe	AC002531	106698 tggagtttctggagt 106712	Y	Alu, hpk
TGGAGCAGTTGGAGC	Ser	AC004474	124745 tggagcagttggagc 124731	Y	EGFR, Smad6
TGGAGTTTGTGGAGT	Leu	U26425	12899 tggagtttgtggagt 12913	PLCb2	PRLR(opposite)
TGGAGCAACTGGAGT	Asn	U96726	61672 tggagcaactggagt 61658	mouse DNA	envelope mIL11R(opposite)
TGGAGTCCCTGGAGC	Pro	U35926	22244 tggagtccctggagc 22230	MHC class II	CFTC, IL6R
TGGAGCAGATGGAGC	Arg	AC002482	14276 tggagcagatggagc 14290	RG208O03	I-309, TcR, IL9R(nonWS)
TGGAGCTCTTGGAGC	Ser	U34879	24914 tggagctcttggagc 24928	EDH17B2	Large tegument protein commonB(opposit, nonWS)
TGGAGCCTTTGGAGC	Leu	Z15025	6359 tggagcctttggagc 6373	Bat2	bat2,mucin, GM-CSFRb(opposite, stop)

Redundant clones were shaded. Highlighted and underlined were Hits and Pseudo-hits respectively.

[Document Name] Abstract

[Abstract]

[Problems to Be Solved] The objective of the present invention is providing novel proteins, and genes encoding the proteins, and uses thereof.

[Means to Solve the Problems] The present invention provides novel hemopoietin receptor proteins, proteins comprising the amino acid sequence of SEQ ID NO: 1 or proteins comprising a modified amino acid sequence of the amino acid sequence of SEQ ID NO: 1 in which one or more amino acids have been deleted, added, and/or replaced with another amino acid, genes encoding the proteins, methods of producing the proteins, as well as uses of the proteins.

[Selected Drawings] None